U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-289 410

Development of a Recommended System for Determining NPDES Method Equivalency

Envirex, Inc., Milwaukee, WI. Environmental Sciences Div.

Prepared for

Environmental Monitoring and Support Lab., Cincinnati, OH.

Dec 78

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA 600/4-78-065	2.	3. RECIPIENT'S ACCESSION NO. PB289410
4. TITLE AND SUBTITLE Development of a Recommended System for Determining NPDES Method Equivalency		5. REPORT DATE December 1978 issuing date
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Envirex Inc. (A Rexnord Company) Environmental Sciences Division Milwaukee, Wisconsin 53201		10. PROGRAM ELEMENT NO. 1 BD 612
		11. CONTRACT/GRANT NO. 68-63-2595
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring & Support Laboratory Office of Research & Development U.S. Environmental Protection Agency Cincinnati, Ohio 45268		13. TYPE OF REPORT AND PERIOD COVERED Final
		14. SPONSORING AGENCY CODE EPA/600/06

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Environmental Protection Agency mandated by PL 92-500 to establish a National Pollutant Discharge Elimination System (NPDES) which involves specification of approved test procedures. This was ammended to include provisions allowing the use of alternate procedures. Both nationwide and limited use (specific permit holder) varinaces are involved. The recommended protocol for nationwide variance evaluation included sampling of five industrial effluent categories and possibly one municipal effluent. Six samples would be collected and analyzed in duplicate by by both the proposed and approved analytical methods. Data evaluation would include a Moments test for normality followed by transformation, if required. The method comparison would involve a pared t-test and regression on the sample means. Data should be analyzed by industrial category and all together in order to identify if the method comparability was limited to certain discharges. Applicants for limited use variances would be required to sample fewer outfalls, however, six samples should be collected which reflect as wide a concentration range as possible. Data analysis is the same as that recommended for a nationwide variance. The recommend statistics were considered applicable for evaluating drinking water variance requests also, and some implementation procedures are included for this system.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Equivalence Statistical Tests Chemical Tests	Analytical Methods Method Comparison	12A

18. DISTRIBUTION STATEMENT Release to Public	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 143
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE PCA07 / MF A01

EPA Form 2220-1 (9-73)

PB 289410

by

R. Huibregtse, N. Kobriger, R. Wullschleger and
Envirex Inc., (A Rexnord Company)
Environmental Sciences Division
Milwaukee, Wisconsin 53201

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

Envirex
a Rexnord Company
Environmental
Sciences Division

DEVELOPMENT OF A RECOMMENDED SYSTEM FOR DETERMINING NPDES METHOD EQUIVALENCY

by

R. J. Fulk, K. R. Huibregtse, N. Kobriger, R. Wullschleger and J. E. Kane
Envirex Inc., (A Rexnord Company)
Environmental Sciences Division
Milwaukee, Wisconsin 53201

Contract No. 68-03-2595

Project Officer

Larry B. Lobring
Office of the Director
Environmental Monitoring & Support Laboratory
Cincinnati, Ohio 45268

ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

DISCLAIMER

This report has been reviewed by the Environmental Monitoring and Support Laboratory and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWORD

To be supplied by EMSL - Cincinnati Staff.

FOREWORD

Environmental measurements are required to determine the quality of ambient and drinking waters and the character of waste effluents. The Environmental Monitoring and Support Laboratory - Cincinnati conducts research to:

- Develop and evaluate techniques to measure the presence and concentration of physical, chemical, and radiological pollutants in water, wastewater, bottom sediments, and solid waste.
- Investigate methods for the concentration, recovery, and identification of viruses, bacteria, and other microbiological organisms in water, and to determine the responses of aquatic organisms to water quality.
- Develop and operate an Agency-wide quality assurance program to assure standardization and quality control of systems for monitoring water and wastewater.

The amendments to the guidelines establishing test procedure for the analysis of pollutants published in the Federal Register, December 1, 1976, specify the methods that are approved for the analysis of pollutants for the purpose of demonstrating that effluent dischargers meet applicable discharge limitations. The National Primary Drinking Water Regulations published in the Federal Register, December 24, 1975, specify the methods that must be used for determining compliance with the requirements of the Safe Drinking Water Act.

Both programs allow for the use of alternate test procedures that have demonstrated equivalency to an approved procedure. This report was developed to compare two procedures and determine if the proposed alternate method is equivalent to the approved method for the particular parameter or contaminant of interest. This report may be implemented as the Agency-wide protocol for methods equivalency.

Dwight G. Ballinger
Director
Environmental Monitoring and Support Laboratory

ABSTRACT

EPA was mandated by PL 92-500 to establish a National Pollutant Discharge Elimination System (NPDES) which involved specification of approved analytical methods. This was amended to include provisions allowing the use of alternate procedures. The objective of this contract was to develop a system for evaluation and comparison of alternate procedures. Both nationwide and specific permit holder variances were considered. The approach involved review of current practices in data handling and statistical methods for outlier rejection, normality testing, transformation, and method comparison. Methods were screened by considering the power of the statistic, limiting assumptions and flexibility, among other variables. The final recommended procedure was chosen based on these results and the application of real data to the proposed techniques. Economic and implementation problems were also considered. The recommended protocol for nationwide variance evaluation included sampling of five industrial effluent categories and possibly one municipal effluent. Six samples would be collected and analyzed in duplicate by both the proposed and approved analytical methods. Then the samples would be mixed, diluted, or spiked to represent six additional concentration levels covering the range of the analytical method. Duplicate analyses are to be performed on these samples. Data evaluation would include a Moments test for normality testing followed by transformation, if required. The method comparison would involve a paired t-test and regression on the sample means. Data should be analyzed by industrial category and all together in order to identify if the method comparability was limited to certain discharges. Single permit holder applicants would be required to sample fewer outfalls, however, six samples should be collected which reflect as wide a concentration range as possible. Data analysis is the same as that recommended for a nationwide variance. The recommended statistics were considered applicable for evaluating National Interim Primary Drinking Water Regulations (NIPDWR) variance requests, and some implementation procedures are included for this system.

After acceptance of the protocol, critical system implementation components were developed. A brief evaluation indicated the feasibility of computerizing the data calculations, and a detailed system specification report was prepared which can be used to develop the actual program. Five types of application forms were developed. First a preliminary form will be sent to the applicant to establish the need for additional data submittal. If necessary, detailed forms will then be sent. Guidelines for program administrations and statistical evaluation have also been presented. The resultant system is ready for use by the EPA.

This report was submitted in fulfillment of Contract No. 68-03-2595 by Envirex Inc. under the sponsorship of the U.S. Environmental Protection Agency. This report covers a period from September, 1977 to July, 1978, and the work was completed as of May, 1978.

CONTENTS

Foreword iii
Abstract iv
Figures vii
Tables viii
Acknowledgements ix

1. Introduction 1
   Project goal 1
   Project objectives 2
   Limitations of the study 2
   Approach to the study 3
   Report format 4

2. Conclusions 7

3. Recommendations 10

4. Review of Current Practices In Data Evaluation 12
   General 12
   Literature review 12
   Some comparative statistical methods used by governmental agencies and professional societies 21
   Direct contacts to government and industrial personnel 25

5. Review of Statistical Methods 33
   Definition of basic statistical requirements 33
   Identification of statistical components 33
   General considerations 42

6. Evaluation of Proposed Methods 44
   General 44
   Development and application of preliminary evaluation criteria 44
   Application of statistical methods to data 46
   Results of statistical tests 54
   Economic considerations 64

7. Implementation Considerations 66
  - Reference methods 66
  - Accuracy testing 67
  - Limited variances 67
  - Type of samples required for NPDES method comparison 68
  - Sample storage system 69
  - Experimental design 69
  - Estimated costs of a variance application 70
  - Data handling 71

8. Development of Recommended Data Evaluation Procedure .. 75
  - Summary of evaluation 75
  - Recommendations for protocol 77
  - Application to the NIPDWR program 79

9. Development of User's Documents 80
  - Introduction 80
  - Development of applicant user's packet 81
  - Development of guidelines for program administration 83
  - Development of guidelines for the review of comparative test data 84

10. References 86

Appendices

A. Literature and data used in review of present practices and evaluation of statistical methods .. 92

B. Feasibility report 133

C. Computer program specification report 141

D. Applications for alternate procedure use-applicants instructions and forms 183

E. Recommended guidelines for administration of alternate method applications 223

F. Reviewers guidelines 228

FIGURES

Number		Page
1	Preliminary Data Handling Approach	34
2	Critical Regions of the Durbin-Watson d-Value	41
3	COD-A Mine Effluent Scatter Plot	55
4	COD-A Inorganic Chemicals Scatter Plot	57
5	COD-A Inorganic Chemicals - Expanded Central Range	58
6	COD-A Inorganic Chemicals - Expanded Extreme Ranges	59
7	COD-B Hospital Scatter Plot	62
8	Overview of the Data Analysis and Storage System	73

TABLES

Number		Page
1	NPDES Specified Parameter By Category	14
2	Statistical Techniques in the Revised Literature	15
3	Comparison of COD Methods on Organic Compounds	17
4	Advantages and Disadvantages of Ourlier Rejection Tests	35
5	Advantages and Disadvantages of Normality Tests	37
6	Advantages and Disadvantages of Non-Parametric Statistical Techniques	39
7	A Comparison of Statistical Results Using COD-A Data	48
8	A Comparison of Statistical Results Using COD-B Data	50
9	Summary of Critical Regions	52
10	Effect of Numbers of Samples and Replicates on Method Comparison Statistics	65
11	Feasible Data Designs	68
12	Estimated Sampling Costs for Equivalency Testing	72

ACKNOWLEDGEMENTS

This report required input from many individuals, and the authors gratefully acknowledge their help. Special thanks are extended to Mr. Larry Lobring, Project Officer, and Mr. Terry Covert for their direction and support throughout the work effort. Ms. Ella Hall and Mr. Paul Britton supplied advice on the statistical evaluation as well as answering questions regarding the data application. Their input was appreciated. Thanks are also extended to other staff members from the Environmental Monitoring and Support Laboratory - Cincinnati, Ohio for their advice and input. The authors also would like to thank those government and industrial personnel who participated in the telephone and in-person interviews conducted during the course of the study. Their candor and helpfulness was greatly appreciated.

Finally, thanks are extended to the staff of the Environmental Sciences Division - technical, administrative and clerical - who participated in this work and contributed to its success. The laboratory staff is gratefully acknowledged for their help in data evaluation and summary of the literature review. Mr. R. Race is thanked for his help in the development of the Detailed Specification Report.

SECTION 1

INTRODUCTION

PROJECT GOAL

The Federal Water Pollution Control Act, PL 92-500, invested the Environmental Protection Agency (EPA) with the authority to limit the discharge of pollutants into a waterway unless such discharges were authorized by the EPA or an EPA approved agency. The resultant program authorized under section 402 of PL 92-500, has been designated the National Pollutant Discharge Elimination System (NPDES) and operates under Title 40, Part 125 of the "Code of Federal Regulations". Monitoring a point source to establish acceptable discharge levels implies a mechanism to select and recommend analytical techniques. Section 304 of the act stated "the Administrator shall, within 180 days of enactment of the title, promulgate guidelines establishing test procedures for the analysis of pollutants that shall include the factors which must be provided in any certification pursuant to section 401 of the act or permit application pursuant to section 402 of this act".

Accordingly, the administrator specified 71 parameters and the accepted analytical techniques for measuring them, which became part of Title 40, Part 136 entitled, "Guidelines for Establishing Test Procedures for the Analysis of Pollutants". The EPA chose to reference established techniques from "Standard Methods for the Examination of Waters and Wastewaters", 13th edition APHA, "Methods for Chemical Analysis of Waters and Wastes", April 1971 EPA and "Annual Book of Standards, Part 23, Water; Atmospheric Analysis", 1972 ASTM.

Title 40, Part 136 was updated December 1, 1976 (Fed. Reg. Vol. 41 No. 232) in response to written comments, suggestions or objectives submitted by Federal and State agencies, major dischargers, trade and professional organizations, analytical instrument manufacturers and analytical service laboratories. Among the modifications approved was the addition of methods from the U.S. Geological Survey and miscellaneous references, the clarification of the conditions of analysis for certain methods, and the inclusion of 44 new parameters.

The intent of Title 40, Part 136 was to provide comparable data throughout the country, thus, enabling the equitable enforcement of the NPDES program. It was recognized, however, that other analytical methods might be suitable for certain applications. Therefore, Title 40, Parts 136.4 and 136.5 allowed applicants to employ alternative methods if such methods are approved by the Director of the appropriate agency. An applicant for an EPA variance must identify the waste stream to be analyzed and provide adequate documentation, in triplicate, justifying the alternate method and establishing its equivalency for use while a final equivalency testing protocol was being developed. The purpose of this project was to develop a protocol for the evaluation of alternate procedures.

PROJECT OBJECTIVES

The basic objectives of the project effort are:

1. To develop and recommend a protocol for statistical evaluation of data generated by alternate and accepted analytical procedures. The results would provide a major criteria for establishing equivalency of NPDES methods.

2. To evaluate various computation systems and develop a program technique for the most feasible system, once a specific protocol has been chosen.

3. To produce a prototype information packet to be sent to individuals requesting a variance. The packet would contain the information necessary for that individual to apply for use of an alternate procedure.

These objectives have been met during the course of this project. The resultant recommended protocol as well as the needed documents for its implementation have been developed. The protocol and the justification for its development and choice have been included in this report along with a discussion of the development of the necessary use documents.

LIMITATIONS OF THE STUDY

It must be understood that a study of this magnitude cannot solve all of the problems which will be encountered when the protocol is implemented. This report does not attempt to provide all of the needed answers to field problems, however, potential difficulties which have been recognized at this time have been highlighted. Where possible, solutions have been proposed, and the statistics have been modified to meet these special needs. Throughout the report, those difficulties relevant to NPDES monitoring and parameter analyses have been presented and the protocol has been chosen to minimize potential problems.

In addition, there is a basic limitation associated with the project goal, since the protocol is mandated by law to establish equivalency, not evaluate the methods in terms of precision and accuracy. Therefore, the statistics chosen were also evaluated with this objective in mind.

Further application of statistics, as well as analyses of known concentrations will be necessary to completely define an analytical technique. Therefore, it should be understood that although this statistical protocol can indicate whether a method is "better" than an NPDES method in terms of precision, the specific comparability of the data involved will be the governing criteria in the overall evaluation.

APPROACH TO THE STUDY

The study was divided into various parts which will be addressed in detail in this report. The first phase involved a state-of-the-art review. Using these results, the second phase included the development, evaluation and recommendation of a proposed method. The third phase involved the development of the various documents needed for implementation of the recommended protocol. This step involved establishing a computation approach and the preparation of a specification report for its implementation. In addition, various application packets were prepared to be sent to those who requested the use of an alternate procedure.

During the state-of-the-art review, those techniques which are presently used in the field were identified, along with problems in method equivalency procedure application. The review involved both a manual and computerized search of the pertinent literature as well as contacts with various personnel including the 10 EPA Regional Quality Assurance (QA) Coordinators. Environmental publications and statistical journals were reviewed. The literature gathered was augmented by applicable sources identified through a computer search. Phone contacts were made with the QA coordinators. Where it was determined to be practical or necessary, trips were made to regional offices for in-person interviews.

After the state-of-the-art review, various statistical methods were identified. Individual procedures to perform the following tasks were identified:

1. Outlier rejection
2. Normality testing
3. Transformation
4. Comparison of mean and variance
5. Evaluation of non-parametric data

The various techniques were presented with regard to their defined advantages and disadvantages. Special consideration was given to statistical power, limiting assumptions, flexibility and ease of calculation. Identification of the most feasible statistical approaches for use in the protocol were evaluated using those criteria. Then the chosen statistics were utilized to evaluate actual data from various sources. Cost considerations were presented along with special NPDES implementation problems. As a result of these steps the final recommendations and justifications were made.

Once the recommended protocol was modified and approved by EPA personnel, some of the necessary implementation documents were developed. First, alternative computational procedures were evaluated in a Feasibility report. Once it was determined that a computerized program was most practical, an anticipated computer flowchart was developed. The information needed for programming the computer was then included in a system specification report which is included as Appendix C of this report.

Concurrently, the needed documents for implementation were being prepared. These documents included applicant user's packets, guidelines for administration of the program and guidelines for the statistical review. The actual documents are also attached as Appendices D-F. However, many of the underlying assumptions involved in their development are included in this report. One important note is that these implementation documents also address applications for alternate procedure requests under the National Interim Primary Drinking Water Regulations (NIPDWR) program. The initial recommended protocol was developed for NPDES method evaluation but it can be used in the NIPDWR system. At the request of the project officer, these additional documents were prepared. However, specific problems associated with NIPDWR have not been considered in detail.

REPORT FORMAT

This report has been presented in sections to show a logical development leading to the recommended protocol and to the development of necessary documents. In an attempt to keep this report relatively brief and to maintain the autonomy of some of the additional documents, a series of Appendices have been attached. It is anticipated that, if needed, certain portions of the Appendix can be reproduced for a specific audience without losing integrity.

The sections and the Appendices are briefly summarized as follows:

SECTION 2: CONCLUSIONS

The conclusions are summarized into the following sections: (1) review of current practices in data evaluation, (2) evaluation of statistical methods, (3) application of statistical methods to data, and (4) implementation considerations.

SECTION 3: RECOMMENDATIONS

The recommendations are itemized and thirteen recommendations are made.

SECTION 4: REVIEW OF CURRENT PRACTICES IN DATA EVALUATION

The results of the state-of-the-art review are summarized. This includes a presentation of the overall literature search results by parameter grouping. Other analytical method evaluation methods now in use are presented as well as a summary of input from personal contacts to state and regional QA coordinators.

SECTION 5: STATISTICAL METHOD REVIEW

This section presents the individual components anticipated to be required in a protocol for establishing equivalency. The various techniques to perform the needed statistical tests are presented and their advantages and disadvantages discussed. A final discussion of basic statistical considerations including hypotheses testing, variance affects and confidence intervals completes the section.

SECTION 6: EVALUATION OF PROPOSED METHODS

The most appropriate statistical tests are chosen based on their advantages with respect to power, flexibility, etc. Several different protocols are identified for further evaluation through the application of real data. A discussion of some of the costs associated with the proposed approaches is included.

SECTION 7: IMPLEMENTATION CONSIDERATIONS

Some of the critical features involved in the implementation of an equivalency protocol are addressed. Input from personnel who must apply the techniques was critical in this discussion. Various features such as the use of a reference method, accuracy testing, and sample handling variables are considered.

SECTION 8: RECOMMENDED PROTOCOL

The results up to this point are briefly summarized, and the recommended protocol is presented. Both nationwide and limited variance requests are considered and the chosen procedures are justified.

SECTION 9: DEVELOPMENT OF APPLICATION GUIDELINES

The background assumptions and basic approach to the development of various application documents is presented. Three sets of information were prepared; application packets, administration guidelines and statistical reviewer guidelines.

APPENDIX A: BACKGROUND INFORMATION

Additional information which supports the state-of-the-art review is included in this appendix.

APPENDIX B: FEASIBILITY REPORT

The necessary changes to the existing Interim Protocol Feasibility report are included.

APPENDIX C: SPECIFICATION REPORT

Information which is required for the development of a computerized data evaluation system has been prepared in this document. Detailed flow charts with alternate input and storage modes are presented.

APPENDIX D: APPLICATION PACKETS

Five different types of application forms have been prepared. These include a preliminary application form plus detailed applications for limited use (case-by-case) or nationwide variance requests under both the NPDES and NIPDWR programs.

APPENDIX E: GUIDELINES FOR ADMINISTRATION

Suggested guidelines for use by Regional Administrators and the Director of EMSL have been prepared.

APPENDIX F: STATISTICAL REVIEWER'S GUIDELINES

Guidelines to aid the statistical reviewer have also been prepared. This includes what a reviewer should look for in the results and some possible interpretations.

SECTION 2

CONCLUSIONS

REVIEW OF CURRENT PRACTICES IN DATA EVALUATION

1. Extensive statistical evaluations or comparisons are not commonly found in literature dealing with chemical methodology.

2. Where information was available, regression and analyses of known concentrations were most commonly utilized for comparison or evaluation.

3. Literature pertinent to radiological analyses was least frequent and metal analyses were most common in the articles reviewed.

4. The EPA Regional and State Quality Assurance Coordinators had a wide range of both interest and expertise in the area of method comparability. All contacted were very helpful. The comments and input provided by the various contacts did not result in widespread agreement. Often suggestions were contradictory. Three comments frequently mentioned were the need for a clear, concise users package, that it should be merged with National Interim Primary Drinking Water Regulations protocol and that the meaning of nationwide variance was unclear.

EVALUATION OF STATISTICAL METHODS

5. The basic screening process involved evaluation of the different techniques based on the power of the statistic, its applicability to distribution types, its ability to handle missing data, its ability to evaluate strata, its general limiting assumptions, the volume of data required, the difficulty in calculation, and common usage.

   As a result of this evaluation, feasible tests included the Moments test and t-test for outlier rejection; the Moments test for normality evaluation; a historical or mean/standard deviation approach to transformation; regression, analysis of variance (ANOVA), or t-test for method comparison tests; Cochran's C test and Bartlett's B test for variance comparison. No non-parametric tests were considered feasible.

APPLICATION OF STATISTICAL METHODS TO DATA

6. Data from two National Pollutant Discharge Elimination System (NPDES) alternate method comparisons were evaluated using the feasible techniques and those included in the EPA Interim Protocol.

7. Throughout all data sets, no outliers were rejected by either the paired t-test or the Moments test.

8. The Moments test proved to be more sensitive in detecting data skewness and kurtosis than the normalized histograms.

9. Log transformation was utilized when data were non-normal. This transform did not always convert the data to a normal distribution. Evaluation of the mean/standard deviation ratio was not effective in identifying the proper transformation.

10. Bartlett's B test was more sensitive than Cochran's C test for establishing homogeneity of variances.

11. The paired t-test most often accepted methods as comparable, especially when one technique was noticeably imprecise.

12. Regression analysis provided the most information regarding the two methods being compared.

13. Regression was adversely affected when the samples evaluated did not cover a wide concentration range.

14. For the data evaluated, violation of the assumption of an errorless independent variable did not seriously affect the results of the regression analysis.

15. The ANOVA was most sensitive in detecting method differences. However, poor analytical precision and missing data points can affect its usefulness.

16. For each of the data sets evaluated, the method comparison tests indicated comparability for some industrial effluent categories and not for others. A method comparability decision based on evaluation of data from all categories would be inappropriate when applied to some of the categories tested.

17. The calculated variance of the analytical methods was shown to greatly affect the final evaluation of comparability.

IMPLEMENTATION CONSIDERATIONS

18. A provision for accuracy testing would greatly increase the information available from method comparability evaluation.

19. Municipal effluent categories can represent a major discharger of many NPDES parameters.

20. Sample mixing, spiking or dilution can allow a greater range of concentrations to be analyzed without adding to the sampling costs.

21. Use of duplicates instead of four replicates does not affect the degrees of freedom in either regression or paired t-tests.

22. The procedure for comparing analytical methods for the NPDES program may be readily adapted to comparing methods for the NIPDWR program.

23. A computerized data evaluation is the most cost effective approach for handling the amount of data expected from this program.

24. The use of two basic types of application forms, preliminary and detailed, will reduce unnecessary data compilation by applicants.

25. Development of four types of detailed application forms will simplify the completion of the report for the applicant.

SECTION 3

RECOMMENDATIONS

1. The number of industrial effluent categories evaluated for a nationwide variance should be limited to five with the provision to add or substitute a municipal effluent. Designation of specific categories should be the responsibility of EPA personnel.

2. The protocol should include provisions for six samples per industrial effluent category. These should be analyzed in duplicate by each method, mixed, diluted or spiked to represent additional concentration levels and reanalyzed in duplicate.

3. The protocol should include evaluation of normality using the Moments test and method comparison using paired t-test and regression.

4. Outlier rejection and variance comparison are not needed and should not be included in the protocol.

5. Evaluation of data sets with different distributions should not be performed. Data from one background population must reflect one distribution in order to be comparable.

6. Non-parametric statistics should not be applied. If data cannot be transformed to reflect a normal distribution, it can be either disregarded or evaluated using normal statistics, as long as the reviewer understands the violation of a limiting assumption.

7. In the case where method comparability is indicated for some industrial effluent categories and not for others, consideration should be given to the concept of issuing variances limited to the appropriate categories.

8. The data evaluation procedure recommended in this report for comparing analytical methods for the NPDES program may also be utilized by Environmental Monitoring and Support Laboratory (EMSL) as an aid in its technical review of NIPDWR variance requests.

9. When evaluating the data, it is necessary to make some comparison of the precision of the two methods. Extreme differences must be considered in the final evaluation.

10. The data evaluation approach should be programmed into an accessible computer system following the recommendations of the System Specification Report.

11. Five types of application packets should be produced. A preliminary application form can be sent to all inquirers and then the appropriate detailed form can be sent when necessary.

12. The basic procedures involved in applying for a variance should be publicized on a nationwide basis, possibly through the Quality Assurance Newletter.

13. The guidelines for administration of the program and review of the statistics should be utilized to aid in consistent implementation of the equivalency programs.

SECTION 4

REVIEW OF CURRENT PRACTICES IN DATA EVALUATION

GENERAL

The first step in developing a viable procedure for establishing equivalency was to review those techniques which are presently in use for comparing analytical methods. This was accomplished by review of 1) the published literature for statistical methods used to compare analytical techniques, 2) the various procedures presently in use for method or accuracy evaluation (American Society of Testing and Materials (ASTM); National Institute of Occupational Safety and Health (NIOSH) Proficiency Analytical Testing (PAT) program; etc.) and 3) the comments offered by state and regional analytical Quality Assurance (QA) Coordinators with regard to implementation of equivalency procedures.

The literature review actually involved three phases. First, articles which presented analytical method comparisons were reviewed and summarized. These results are presented in this section of the report. The second aspect of the literature review involved a study of statistical techniques to establish the advantages and disadvantages of each technique. These results are presented in Section 5. The final aspect of the review was an evaluation of established techniques presently used to compare analytical methods or laboratory results. In this portion of the study, the approaches used by various governmental and private groups for establishing approved methods were reviewed. These procedures are summarized and briefly discussed with regard to their applicability to this project.

The personal contacts with people involved in equivalency programs included initial telephone contacts followed by visits to the individuals particularly active in equivalency procedures. For the sake of clarity, all input has been summarized by subject only.

LITERATURE REVIEW

Introduction

As indicated, the purpose of this review was to investigate the statistical methods being used by authors whose work compared analytical techniques or investigated interlaboratory variances by collaborative studies. The object was to determine if some parameters required special statistical applications or if insights into data development in general would be useful in establishing an NPDES alternate method protocol.

The review was generally limited to articles published within the last ten years. Special emphasis was placed on those published after promulgation of the NPDES approved methods (40 CFR 136, Oct., 1973). The papers were separated into categories and reviewed. Priority was given to those authors who used NPDES approved methods (see Table 1 for categorical listings), although papers were not excluded if they compared non-approved techniques or were concerned with analytical methods used in non-environmental disciplines. The overriding consideration was that comparisons were being made between two or more techniques, laboratories or sample sources.

Many fine articles were not included because they used non-comparative techniques, or they offered no statistical arguments in their data comparison Within the context of this review, the term basic statistics applies to mean and standard deviation, and includes related terms such as coefficient of variation, range and variance. A formal statistical argument may be defined as an experimental design, or structuring of the data such that hypotheses may be tested against predetermined critical values.

Most of the literature search was done manually. The Annual Reviews published by the Journal of the Water Pollution Control Federation and Analytical Chemistry provided the basis for the manual search. Consultation with Envirex's staff librarian and EMSL-Cincinnati's librarian indicated a keyword search would be of limited value since statistics which are used as data handling tools are often not included in the abstract. Nevertheless, approximately 700 environmentally related abstracts supplied by the University of Wisconsin Library, through a limited computer search, were reviewed and additional pertinent articles were obtained.

A summary of the literature review has been given in Table 2. The detailed information is presented by review category in the Appendix (Table A-1). The literature review in this section is discussed by analyte category and summarizes the methods of handling data distributions and transforms, dealing with samples of known concentrations, and the formal statistics used in data comparison. Papers not cited as using formal statistical arguments used basic statistics (mean, standard deviation) and compared the results in qualitative terms.

Bacteriological

The ten papers in this category dealt primarily with comparing membrane filter (MF) techniques with the five tube most probable number (MPN) technique. This category had the largest number of papers which used a log transform before data evaluation. Sladek (1) and Schaeffer et. al. (2) argue that the log transform is necessary because bacteria suspended in water tend to follow the Poisson distribution, which requires a log transform to normal (i.e., arithmetic) error distribution. The implications of applying non-normal distributions to conventional statistics will be discussed in detail in a separate section.

TABLE 1. NPDES SPECIFIED PARAMETERS BY CATEGORY

(Parameter name followed by NPDES parameter number)

Bacteriological					
Fecal coliform	4,5	Fecal Streptococcus	8		
Total coliform	6,7				
Demand					
BOD	10	COD	12	TOC	91
Metals					
Total and soluble species	24 - 87				
Nutrients					
TKN	23	NO_3	88		
Organic N	92	Total P	97		
NH_3	3	Elemental P	97		
NO_2	89	Ortho P	93		
Organics					
Benzidine	9	Pentachlorophenol	94		
Chlorinated organics -	14	Pesticides	95		
Cyanide	17, 18	Phenols	96		
Oil & Grease	90	Surfactants	113		
Physical, Mineral, Residue					
Acidity	1	Total Dissolved Solids	105		
Alkalinity	2	Suspended Solids	106		
Chloride	13	Settleable Solids	107		
Chlorine	15	Total Volatile Solids	108		
Color	16	Sp. Conductance	109		
D.O.	19	Sulfate	110		
Fluoride	20	Sulfide	111		
Hardness	21	Sulfite	112		
pH	22	Temperature	114		
Total Solids	104	Turbidity	115		
Radiological					
Alpha - Total	99	Beta-count error	102		
Alpha-Count error	100	Radium	103		
Beta-total	101				

TABLE 2. STATISTICAL TECHNIQUES IN THE REVIEWED LITERATURE

Analyte category	Papers reviewed	Papers using log transformed data	Papers challenging by samples of known concentration	Papers comparing data with formal statistics
Bacteriological	10	6	0	8
Demands	7	1	7	4
Metals	20	3	16	9
Nutrients	14	1	9	8
Organics	11	0	9	2
Physical, Mineral, Residue	19	0	8	8
Radiological	3	0	1	0
Total	84	11	50	39
%		13%	60%	46%

Statistical Technique Employed

Rejecting Outliers	Number of papers
t-test	5
Chavenet's criterion	2
Dixon's	1
Evaluating Distribution	
histograms	2
skewness	2
Comparing Data Populations	
Regression	21
t-test	12
F-test	6
ANOVA	5
Youden's Pairs	3
Duncan's Multiple Range	1
Chi Square	1

This is the only category which used no spiked samples or samples of known concentration to evaluate accuracy. There is presently no acceptable technique for preparing a primary standard microbiological sample. Authors generally attributed the most accurate results to the MPN technique.

The most diverse and sophisticated formal statistical experimental designs were used in this category. Eight papers (80%) employed more than basic statistics. The regression (least squares curve fit) was used most frequently. (2,3,4,5,6).

Schaeffer et.al. (2) cautioned against the use of a model I regression which assumed no error in the X variable and recommended a model II regression which allowed errors in both variables being evaluated. Four authors who used the regression analysis further evaluated the confidence intervals of the slope (4), or of the slope and intercept (2) or evaluated the residuals by means of an analysis of variance (ANOVA) (5) or t-test (6). The significance of thoroughly evaluating the regression equation will be discussed with examples in the nutrient category. Other statistical comparisons were conducted by ANOVA (7) and Duncan's Multiple Range Test (8). Duncan's method is an elegant technique for comparing several means at once. However, it has limited application to NPDES alternate method evaluation because there would rarely be a need to compare more than two methods.

Demand

Authors comparing methods for Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), and Total Organic Carbon (TOC) did not generally feel a log transform was necessary before evaluating the data. Only Verstraete, et.al. (9) used a log transform when comparing a manometric BOD to the conventional closed bottle technique. Winter, et.al.(10) evaluated the distribution of TOC, BOD and COD data gathered in an EPA interlaboratory study by means of skewness statistics (which measure curve symmetry) and histograms but concluded no transforms would be necessary.

All papers reviewed in this category challenged both analytical methods with samples of known concentrations. The importance of using samples of independently defined concentration may be best demonstrated by the work of Jirka, et.al. (11). This paper compared the NPDES approved reflux COD with a semi-automated technique which involved closed tube digestion. Thirteen samples containing known concentrations of organic materials were evaluated (Table 3) using a paired t-test. The results reported for pyridine and 3-picoline were rejected as outliers using a t-test.

The results of the paired t-test showed the standard method and the candidate method were not equal (95% confidence interval). In the absence of "known concentration" information, the conclusion was that the methods were not comparable. However, when the methods were compared to the theoretical yield, it was clear that the standard method tended to underestimate the available COD. In this instance, using samples of known concentration provided valuable information about method accuracy.

TABLE 3. COMPARISON OF COD METHODS ON ORGANIC COMPOUNDS

(from Jirka et. al. (11))

Compound	COD, mg/1 Theoretical	Standard method	Semi-automated method
Phenol	238	230	240
Sodium acetate	470	450	462
Acetate	221	200	207
Ethanol	197	170	180
Dextrose	107	100	113
Methanol	150	130	139
Oxalic acid	127	120	128
Sodium citrate	490	470	496
Glutamic acid	98	98	98
Glucine	64	60	62
Benzoic acid	197	190	202
Pyridine*	223	5	3
3-Picoline*	239	64	77
Tetrahydrofuran	244	250	242

* rejected by t-test as outliers

Comparison	t-value	Critical t-value (95% confidence)	Decision
Standard to semi-auto	-3.60	±2.201	not equal
Standard to theoretical	-3.87		not equal
Theoretical to semi-auto	1.24		equal

Four authors used some form of formal statistics to compare their data. Verstraete (9) and Tebbutt (12) fitted their BOD data to a least squares regression but did not evaluate equation confidence intervals or residuals. Tuffey, et.al.(13) used an F-test to evaluate the variance homogeneity between long term manometric and closed bottle BODs. Winter (10) evaluated the accuracy (as % relative bias), 95% confidence intervals and relative deviation (coefficiant of variation) of various laboratories in a collaborative study of BOD, COD and TOC methods. As in the Bacteriological category, regression analyses were used most often but, in general, fewer and less sophisticated statistical designs were applied to Demand data.

Metals

More papers were reviewed for this category than any other. Three authors out of 20 determined that a log transform should precede data evaluation. Salotto, et.al. (14) evaluated the concentrations of metals in US waste treatment plant sludges. An examination of the normalized histograms led the authors to transform the data to minimize skewness. Yeager, et.al. (15) found that the concentration of lead in soils followed a log normal distribution. Jackson (16) reported similar results for cadmium in soil. It is perhaps significant that no workers investigating metals in the water or wastewater environment transformed data before statistical evaluation.

Sixteen papers (80%) reported using samples of known concentration or spiking to evaluate method accuracy. Only the categories of Demand and Organics had higher percentages of this type evaluation. This finding may indicate that there is a preference to evaluate methods with samples of known concentration when the analyte is available, as it is for metals.

Nine authors employed formal statistical experimental designs in data comparisons. Regressions were used most often; four authors chose this technique to evaluate the data. Raman (17) reported the significance of the regression coefficient (r) for the concentration of cadmium in biological serums using the Murexide and ion specific electrode techniques. Yeager, et.al. (15) used the 95% confidence intervals of the regression slope to comment on the goodness of fit to two atomic absorption techniques for determining lead in biological serums. Ostling (18) and Jackson (16) presented the least squares curve with no further data evaluation. Lenz, et.al.(19) demonstrated the usefulness of the ANOVA in residual analysis when the effect of different paper mill wastewater flows on four different methods of sodium analysis was evaluated. Their blocking approach for the two way ANOVA could be useful in evaluating the effect of different effluent sources on the approved and candidate methods in a NPDES variance application. Winter, et.al.(2) used Youden's nonreplicate sample method to evaluate the effect of inter-, and intra-laboratory variances on data from the cold vapor mercury technique. This design uses Youden squares and the F-test to evaluate variance. While it is an elegant and useful technique for interlaboratory collaborative studies, it requires samples of known concentration to be effective. It would therefore be of limited value in evaluating methods with NPDES monitored outfalls of varying and unknown concentrations. Robbin, et.al. (21) evaluated the data from a collaborative study of arsenic-in-air methods using the Chi-squared technique to test goodness of fit.

In summary, the metals category was characterized by few transforms, many challenges by known concentration samples and statistical designs of moderate sophistication.

Nutrients

Nutrient parameters together with Residues and Demand Parameters, account for most of the monitored analytes in the NPDES program. Of the fouteen papers reviewed, only Yu, et.al. (22) transformed the data before analysis although Winter, et.al. (23) evaluated the skewness of the data gathered in an interlaboratory collaborative study and found it to be normally distributed. Outliers were rejected by the t-test (23) and Dixon's test (24).

Spiked recoveries and standard "known" concentrations were reported in 9 papers (64%). Jirka, et.al. (25) used spike recoveries to indicate a potential negative bias in the NPDES approved total phosphorous method. Spiked recoveries were also used in evaluating interference effects (22, 25, 26, 27, 28), as well as bias tendency.

Regression was the most frequently used evaluation technique. Eight authors chose formal statistical techniques for data comparison, of which six were regression. The previously cited work of Jirka, et.al. (25) illustrates the importance of complete regression equation evaluation. The regression equation will be dicussed in detail in following Sections. However, for purposes of this discussion, a brief review is appropriate here. The linear regression equation has the form given in equation (1).

$$\text{Equation (1)} \quad Y = b_o + b_1 x + e_1$$

where: Y = response or dependent variable
x = independent variable (assumed to be known without error)
b_o = intercept
b_1 = slope
e_1 = error

Confidence intervals may be calculated for each term in the equation. The regression coefficient, r^2, is calculated from the sum of squares (S.S.) due to regression divided by the sum of squares about the mean (i.e. S.S. about regression + S.S. due to regression). The greater the variance accounted for by the regression, the better the fit and the larger the r^2 value (r^2 has a maximum value of 1). The related r statistic, the correlation coefficient, is most often cited in regression studies. An r value which approaches ± 1.0 signifies a "good fit" regression. However, without adequate qualification, it can be misleading.

Jirka, et.al. evaluated a micro total phosphorous procedure against the NPDES approved method using 35 industrial and municipal effluents (25). The regression yielded a slope of 1.05 and an r value of 0.997. Without further evaluation, the two methods would appear to be comparable. However, the authors evaluated the means of the methods using a t-test (95% confidence interval) and found the micro technique was significantly different (high bias) from the standard method. This example illustrates that misleading information can result from incomplete data evaluation.

Others who used regression in this section reported the significance of the r value (29), gave the confidence interval of the slope (26) or simply reported the best fit line in graph form (24, 30, 31). Winter and Midgett used Youden's design for the EPA collaborative study of nutrients (23) and Yu, et.al.(22) used the ANOVA when comparing the brucine nitrate method to the ion specific electrode.

Authors who did comparative studies in nutrient analyses: 1) generally did not transform the data before statistical evaluation and 2) utilized the regression statistical design.

Organics

The methods included in the organic category tend to be more complex and expensive to perform. This may account for the fewer comparative studies reviewed for organics. None of the authors used transforms prior to data analysis and no data distribution evaluations were reported. Researchers comparing organic methods made extensive use of known concentration samples. One application was a study of preservation techniques of the phenol procedure by Carter, et.al.(32). In this study, the workers used spiked samples to evaluate the efficiency of standard and modified preservation methods with time. Nine authors (82%) used either spiked recoveries or known concentration standard to evaluate their methods (32, 33, 34, 35, 36, 37, 38, 40).

This category was unique because of its extensive use of known concentration challenges, as well as the fact that none of the reviewed papers used the regression model for data evaluation. Only two papers used formal method comparison techniques. Meyer, et.al.(33) used the t-test and the F-test in comparing the data generated from an interlaboratory study of refinery wastewater. Carter, et.al.(32) used the t-test in the evaluation of phenol preservation methods. The rest of the studies presented the basic statistics with no further development.

Physical, Mineral, Residue

Although this was the second largest category in terms of volume (19 papers), no log transforms were reported and none of the authors reported any formal evaluation of the data distribution. Eight papers (42%) used samples of known concentration to challenge method accuracy (41, 42, 43, 44, 45, 46, 47, 48).

Formal statistical design techniques were used for data evaluation in eight papers; four of them were regression techniques. ANOVA tables were used by Finger, et.al.(49) to evaluate the regression residuals from an automated spectrophotometric suspended solids technique and the NPDES approved method.

The 95% confidence intervals for the slope of the regression comparing an ion specific electrode technique with the SPADNS method for flouride was reported by McLeod (45). Others reported the regression equation with no residual evaluation (50, 51). Winter, et.al.(41) used Youden's technique in evaluations of physical, mineral and residue interlaboratory variances.

Radiological

The radiological category had the fewest papers for review in this study. None of the papers in this category transformed the data before evaluation or used formal statistics in data development. Kanipe (52), in his handbook for quality control in radiological laboratories, recommended the frequent use of known concentration standards as method challenges. He also discussed the reasons for the Poisson distribution of radioactive materials in the natural water matrix. The three articles reviewed at this writing were not considered adequate to determine data handling trends in this category.

SOME COMPARATIVE STATISTICAL METHODS USED BY GOVERNMENTAL AGENCIES AND PROFESSIONAL SOCIETIES

This section reviewed data comparison procedures used by some governmental agencies and professional societies. The governmental programs which were studied included the EPA Interim Protocol, EPA interlaboratory studies, the EPA Ambient Air Quality Reference and Equivalent Methods Protocols, and the National Institute of Occupational Safety and Healty (NIOSH) Proficiency Analytical Testing (PAT) program. The American Society of Testing and Materials (ASTM) and the Association of Official Analytical Chemists (AOAC) were the professional societies whose programs were reviewed.

The EPA air quality monitoring variance program, and the NIOSH PAT program are structured, on-going projects which provide information pertinent to implementing and upgrading techniques for data comparison. The ASTM and AOAC recommended statistical techniques, as well as the EPA interlaboratory collaborative studies, are a compilation of many statistical methods. Their value to this review was their discussions concerning the techniques which might be most suitable in providing information about comparative NPDES analytical testing.

EPA Interim Protocol

The EPA Interim Protocol is used to evaluate the comparability of alternate analytical techniques to NPDES approved methods (40 CFR 136). Applications for a permit holder or a nationwide variance are possible. These applications are forwarded through the appropriate state officials to the Regional Administrator. Typically, the Regional Quality Assurance Coordinator will review the data and forward any recommendations and the data to the Director, EMSL-Cincinnati. The EMSL-Cincinnati staff members then perform additional technical review, statistically evaluate the data and forward their conclusions to the Director for final recommendation.

Variance applicants for nationwide approval are requested by EPA to sample the five major industrial effluent categories in which the analyte is monitored. Six concentrations are to be taken from each discharge and replicated four times using both the proposed and approved methods. A nationwide variance requires 240 individual analyses. Outliers are rejected by a two tailed t-test and the distribution of each industrial effluent set evaluated using a seven cell normalized histogram. Variance homogeneity is evaluated by Cochran's method and data is stratified if necessary. Mean equality is evaluated by the t-test for common strata. The statistical design is simple enough to be understood by applicants and yet does not assume that the limiting assumptions of the test statistics are automatically met.

EPA Interlaboratory Testing Program

One element of the EPA's laboratory quality assurance system is the interlaboratory testing program. The EPA has conducted a number of interlaboratory collaborative method studies, some of which have been previously cited. Statistical development in these studies range from relatively simple data presentations in the Demand research study(10) to the most sophisticated applications in the Mineral and Physical study (41), the Manual Nutrient Method study (23) and the Cold Vapor Mercury study (20). Common to all these studies were techniques to reject outliers and the use of known concentration samples to evaluate bias. Three of the studies evaluated data distributions (10, 23, 41) by means of histograms and skewness statistics, and three used Youden's squares for inter-, and intra-laboratory variance evaluation (20, 23, 41).

Green, et.al. (53) in their state-of-the-art review for conducting an EPA interlaboratory testing program had two recommendations which were applicable to the NPDES alternate method protocol: 1.) "The statistical analysis should be as complete as possible so that accurate program results may be reported to the participants" and 2.) "In order to standardize proficiency evaluation, the statistical analysis procedure adopted by the EPA should be published in 'User's Manual' format for use by any inter-laboratory test program manager involved in water quality measurements".

EPA Ambient Air Quality Program

This program has implemented a "user's package" for the evaluation of alternative analytical methods proposed for Reference or Equivalent Method Status in Air Quality Monitoring. The regulations for method approval have been promulgated in 40 CFR part 50 and part 53 under the title "Ambient Air Monitoring Reference and Equivalent Methods." These regulations clearly define the amount of data and test conditions required to establish comparability to a reference method. By specifying a reference method of high precision and accuracy, the EPA assures continuity in the air quality monitoring program while minimizing the potential of admitting mediocre methods. The technical community is notified of newly approved methods through the Federal Register and the EPA quarterly Quality Assurance Newsletter. A concise, easily understood user's package, and notification of newly approved methods to the NPDES monitoring community, would be desirable elements in an NPDES alternate method protocol.

NIOSH Proficiency Analytical Testing

The NIOSH Proficiency Analytical Testing (PAT) is similar to the EPA Interlaboratory testing program in that its primary objective is to monitor performance of laboratories under contract to NIOSH. Participants' analytical abilities in measuring metals, asbestos and organics are tested. Samples of known concentration are sent to participating laboratories and the results are returned on NIOSH-PAT report forms. Four concentration levels are analyzed for each parameter. The statistical protocol is unique in that it assumes log normal distribution of the data (based on previous experience), and transforms all data before statistical evaluation. The program is capable of evaluating inter- and intra-laboratory variation.

The critical test regions for intra-laboratory variation are determined by evaluating the residuals (residual=observed value minus the filter effect minus the laboratory effect minus the grand mean) using a two way ANOVA. The maximum allowable residual is calculated after the method of Anscombe and Tukey (54). Values which exceed the critical value are rejected and the evaluation of residuals is repeated until no rejections occur. The mean of the remaining values is the predicted value. The intralaboratory variance is then reported as a ratio of observed/predicted values for each of the four concentrations of a given analyte. The results are reported in a computer printout showing the critical limits and the scatter of the observed values for the reporting laboratory. Interlaboratory variance (systematic error) is evaluated in a similar manner. Laboratories with excessive intra-laboratory variation are eliminated from interlaboratory calculations. Critical regions are given by the t-distribution at the 95% confidence interval. Extreme values are eliminated from the grand mean until the t-value falls within the critical region. Each laboratory mean is reported as a ratio of that mean to the grand mean.

This statistical protocol is unique in its use of log transforms and its extensive evaluation of residuals. Although the statistical manipulations are complex, the results are easily interpreted in graphical form. It is unlikely an NPDES recommended protocol could assume all data to have a given distribution, in light of the many analytes for which NPDES approved methods exist. However, evaluation of residuals against critical regions would provide useful information in determining method comparability.

ASTM and AOAC Programs

The American Society of Testing Materials (ASTM) and the Association of Analytical Chemists (AOAC) are voluntary professional societies with analytical method evaluation programs. Because of the inter-disciplinary scope of these societies, they do not propose rigid statistical formats for method comparison. They do provide recommendations for performing comparative studies which are widely followed. This discussion will center on the parts of their programs which would be useful in providing comparability information for the NPDES alternate method protocol.

ASTM Procedures--

The ASTM D-19 committee on water has the responsibility for conducting collaborative tests to evaluate new analytical methods prior to inclusion in the ASTM Part 31 Methods Manual (55). Forty-one ASTM methods have been cited as approved for NPDES monitoring (40 CFR 136). If there is interest in a specific topic, such as an improved method for mercury analysis, a task group can be formed to investigate the method. A results advisor is appointed to help design the collaborative study and interpret the results. The recommended protocol for a collaborative study is given in the ASTM Methods Manual, Part 31, under designation D-2777. The statistical format for determining method precision and bias contains four major components. Protocol recommendations are given below.

1. Preliminary Studies: Pilot work should evaluate variables such as reaction time, temperature, reagent concentrations, interferences, calibration, and sample size. Youden's ruggedness test is given as an example (A-5).

2. Outlier Rejection:
   Outliers Laboratories - Youden's ranking techniques (A-5)
   Outlier Observations - t-test for extreme values

3. Precision:
   Overall Precision: If precision is constant over the evaluated concentration range.

$$S_T = \sqrt{\frac{(n_{L1} \times S_{L1}^2) + (n_{L2} \times S_{L2}^2) + (n_{L3} \times S_{L3}^2)}{n_{L1} + n_{L2} + n_{L3}}}$$

   if precision varies with concentration, plot S_L verses X_L and report precision by concentration. If possible, fit the curve using a least squares regression and report the resulting equation.

4. Bias:
   Bias = average amount found - amount added.
   Test the significance of the residuals with a t-test.

In practice, the task group may choose more sophisticated experimental designs as in the mercury cold vapor collaborative study previously cited (2). Basically, the ASTM D-19 committee feels an effective method study should include the following components: 1) a prestudy pilot study (conducted by the applicant), 2) a technique for rejecting outliers, 3) a statistical statement of method precision (by concentration if necessary) and 4) a way to measure method bias against known concentrations. Because the ASTM D-19 committee does not require comparative testing, their recommended statistical format is not directly applicable in developing a recommended protocol. One ASTM technique which is valid for a technological comparative study is the peer review system. Candidate methods are subjected to technical as well as statistical evaluation. Reviewer comments must be answered to the satisfaction of the commenter or the task group.

AOAC Comparative Method--

Tha AOCA is similar to ASTM in many ways. It is composed of members who are analytical scientists in government regulatory agencies and associate members who are non-government scientists. Analytical methods adopted by the AOAC have been validated by collaborative study and technical review and may be found in the "Official Methods of Analysis of the AOAC, 12th edition. (1975)." Annual supplements contain method revisions and newly developed methods. Twenty-three AOAC analytical procedures have been approved for NPDES monitoring.

The AOAC has no formal statistical protocol for evaluating candidate methods. If a collaborative study is needed, the AOAC Executive Office assigns the topic to a General Referee who in turn appoints an Associate Referee to undertake the study. The Associate Referee is responsible for designing and evaluating the collaborative study. A review of AOAC recommended statistical techniques has been published by Youden and Steiner (56). It is the stated policy of the AOAC to conduct these collaborative studies for the purpose of evaluating the method under actual working conditions.

The AOAC statistical design often used in interlaboratory studies is Youden's non-replicate pairs. Youden attempted to keep the statistics simple and the results easy to interpret. The basis of Youden's design is the non-replicate, similar sample technique which has been used in three reviewed papers (20, 23, 41). The design is simple and the graphic data presentation is easily interpreted. The key to Youden's non-replicate pairs is a source of similar, non-replicate samples. It would be difficult to predetermine sample concentrations in NPDES monitored outfalls. This would render Youden's technique inappropriate for NPDES method evaluation.

DIRECT CONTACTS TO GOVERNMENT AND INDUSTRIAL PERSONNEL

General

In order to augment the information obtained through the literature search, direct personal contacts were also made to government and industrial personnel who had an interest in the equivalency procedure. These contacts included the 10 EPA QA coordinators, several state agency personnel and interested industrial people. Initial contacts were made by telephone and after evaluation of these results, trips were scheduled to those agencies which seemed able to provide the most insight. In addition to these meetings, an initial project "kick-off" meeting was held in Cincinnati. At this meeting, several EPA-EMSL personnel offered their comments and interpretations of the equivalency procedures.

This section presents a brief summary of the information obtained from these meetings and telephone conversations. The extent of the responses varied directly with the individual's experience in handling or working with variances. During the discussions, comments about the equivalency process were made that concerned matters outside the scope of the project. However, they are included in this report in the spirit of openness with which they were given.

Initial Meeting in Cincinnati

Important information was gained at the kick-off meeting held in Cincinnati at the EPA offices there. Comments were obtained from the following people:

Mr. Larry Lobring	Chemist EMSL - Cincinnati
Mr. Terry Covert	Microbiologist EMSL - Cincinnati
Ms. Ella Hall	Statistician EMSL - Cincinnati
Mr. Dwight Ballinger	Director EMSL - Cincinnati
Mr. Robert Booth	Deputy Director EMSL - Cincinnati
Mr. Paul Britton	Statistician EMSL - Cincinnati
Mr. John Winter	Branch Chief EMSL - Cincinnati

The most important input received was a copy of the EPA Interim Protocol presently being used in evaluation of variance requests. This procedure, developed mainly by Mr. Britton, allowed an initial starting point for technique evaluations.

Several important aspects of the project were stressed. Among the points discussed were the following:

1. Although the NPDES method may not be best, it is the approved technique by law and comparability to this method is the object of the statistical evaluation.
2. The project goal was to develop a workable program format for direct application by EPA.
3. The range of method evaluation considered was agreed to be the range of the analytical method due to difficulty in establishing pertinent ranges for NPDES parameters (since permits are written in weight per time units).
4. Spiked samples for evaluation of bias would be advantageous.
5. Strata effects should be evaluated in the protocol.
6. Consideration should be given to evaluation of the experimental variance since excessive variance could bias the final results.

These comments were taken into consideration throughout the project performance period.

Contacts to State and Regional Personnel

One of the most important aspects of this survey was to contact people, primarily Quality Assurance (QA) Control Coordinators, who have direct contact with applicants requesting permission to use an alternate procedure. In two instances, the regional coordinator recommended that state officials be contacted for additional input.

All of the ten regional QA coordinators plus the State of California were contacted initially by telephone and asked the questions given in Appendix Figure A-1. The results of these conversations were evaluated and trips were scheduled to those facilities which could provide the most insight into the NPDES Alternate Method Program or showed a special interest. At the recommendation of Mr. David Payne, additional trips were also scheduled to the States of Michigan and Minnesota.

The agencies visited and the people interviewed are listed below:

1. Region 8 EPA and National Enforcement Investigation Center (NEIC) - EPA, Denver, Colorado:

Dr. Mark Carter	Deputy Chief, Chemistry Branch, NEIC
Dr. Theodore Meigs	Chief, Chemistry Branch, NEIC
Mr. John Tilstra	Chief, Chemistry Section; Technical Services Branch

2. Region 5 EPA, Chicago, Illinois:

Mr. David Payne	QA Coordinator

3. Region 4 EPA, Athens, Georgia:

Mr. Bobby Joe Carroll	QA Coordinator
Mr. Tom Bennett	Section Chief

4. EMSL - Dept. E. Research Triangle Park, North Carolina:

Mr. Larry Purdue	Chief, Monitoring Techniques Evaluation Section; Quality Assurance Branch, EMSL

5. Region 2 EPA, Edison, N.J.:

Mr. Gerald McKenna	QA Coordinator

6. State of Minnesota, St. Paul, Minnesota:

Mr. Al Dressler	Twin Cities Round Robin
Mr. John Davenport	MPCA
Mr. Russel Frazier	MHD
Mr. Keith Hoffmann	Serco Laboratories

7. State of Michigan, Grand Rapids, Michigan:

Mr. T.K. Wu	Michigan Environmental Services Division Laboratory Director
Mr. Polasek	Michigan Department of Natural Resources Permit Section

The following paragraphs summarize the results of these discussions.

General Comments--
The purpose of the telephone contacts and meetings was to gather information about past experiences with the equivalency process. The state, regional, and NEIC-Denver QA coordinators were helpful and direct in responding to questions about the present variance procedures. In some cases, the regional QA coordinators provided contacts at the private or state level which were useful in obtaining insights on data handling, and provided suggestions and comments pertaining to the NPDES alternate procedure program.

Comments About the Present Protocol--
The regions generally had more experience with single-source, permit holder discharge variances than the laboratory-wide, state-wide, region-wide or nation-wide variances. The number of variance applications processed per region ranged from 3 to more than 40. Most variances were for nutrients (NO_3-N, NO_2-N, TKN, phosphorous) or metals. Fewer applications were reported for BOD_5, residue, organics or microbiological methods and none for radiological methods. Some of the specific variances which have been granted include the use of a Warburg Manometer for BOD_5 measurements; a Hach Manometer for BOD_5 measurement; an infra-red spectrophotometer for oil and grease analysis; a GC/MS for pesticide and PCB analyses; and the use of a modified COD test involving closed cell digestion and colorimetric measurement of dichromate concentration.

State experiences were also varied. State method variances were limited to permit holder type. One of the states commented that although they receive many initial requests for variances, no applications have been processed. It was felt that potential applicants were discouraged from applying because: (1) the instructions were too vague, (2) it appeared that the required testing was too extensive, and (3) it was expected that a long time would pass (say one year or more) before the applicant would be granted a variance.

State and regional coordinators generally followed the data format recommended by the EMSL-Cincinnati Director. Most of the QA coordinators reviewed the data before forwarding it to EMSL-Cincinnati. If the data appeared to be unacceptable, the request was returned to the applicant with the suggestion they locate and correct the source of poor data before resubmitting the application. Two of the QA coordinators used fundamental statistics to evaluate the data. The data was challenged if: 1) the means were more than three standard deviations apart, based on the variance of the NPDES approved method, 2) the proposed method was not within ± 10% of the approved method or 3) the results from the proposed method were consistently lower than the approved method. It was estimated that spiked recoveries were done on about half of all the point source variances from all regions. These were most often used to resolve apparent method inconsistencies and most coordinators relied on EMSL-Cincinnati for direction in terms of when to require spiked samples.

One state procedure involved evaluation of the proposed technique prior to any testing. This included consideration of the type of waste as well as the proposed procedure. Test data is first submitted to the State where, if approved, it is submitted to the Region. The statistics used in data evaluation at the state level were not defined.

There was much confusion among the QA Coordinators about the scope of variances beyond the point source category. Although regional, state-wide and laboratory variances have been granted, most of the laboratories would like to see the issuance of only permit-holder and nation-wide variances.

Two laboratories suggested that variances granted to regional EPA laboratories have automatic nation-wide status to eliminate the duplication of all the test data by other regional labs that wish to use the alternate procedure. Four of the regions reported handling either laboratory-wide, state-wide, or region-wide variances; however, in three of the four regions questions about number of samples and sample sources were directed to EMSL-Cincinnati. There were questions about authority to grant variances and what effect a region-wide variance would have on the nation-wide scope of the NPDES program.

One commenter indicated that variances should be limited to specific concentration ranges. He indicated that variances applicable to one industry or to one industrial effluent category would be preferable to nation-wide variances.

The regional laboratories want to be sure that they retain a role in the equivalency process. One laboratory mentioned that they prefer to act only as the middleman between the state and EMSL-Cincinnati and do not want to do any data evaluation themselves. They want to be familiar with the statistics, however, to help them in their discussions with the applicants. This laboratory also suggested that one central laboratory (possibly at Cincinnati) do the analytical testing for nation-wide variance applications. Other laboratories, however, prefer that the testing be performed by the applicant. State personnel felt they sould be involved in the variance evaluations. It was suggested that evaluation be handled in stages with state agencies handling in-state applications.

Suggestions for the Proposed Protocol--

The following comments were made in response to the question, "What is your assessment of the advantages and disadvantages of the method you use?" and "How would you make the method better?" Five of the regions surveyed felt a standardized "user package" which contained clear and concise instructions on applying for an alternate procedure would be useful. Six of the regions felt that any "user's package" should contain information on how the data was to be evaluated. One region felt the standard form might limit the flexibility of the program while another felt the applicant should not know the statistics being used. One group suggested that an equivalency testing procedure include the following:

1) clear and concise instructions, including the exact number of samples to test, the number of analyses to perform and the data required by the EPA for evaluation purposes,

2) a definite time in which the applicant might expect an answer. A period of 60 to 90 days was considered reasonable,

3) an expert in the EPA who would be readily available to the applicant to answer questions concerning the equivalency testing and, once the data is submitted, be able to inform the applicant of the status of his application.

The laboratories disagreed as to the type of equivalency comparison method that should be used. Two commenters thought that a very specific procedure, complete in all details, should be prepared and published as regulations so that the applicants know exactly what is required. One laboratory, however, wants a very flexible procedure that will allow the applicant to set up the experimental design he feels best suited to his situation. Another laboratory would like to have the method flexible enough so that the applicant could use data obtained in the past as part of his application.

Three regions wanted one or more of the approved methods for a given parameter to be a reference method for the purpose of equivalency testing, similar to the Air Regulations 40 CFR 50, 53. Using this approach, all tests would be compared to the same method. Another commenter suggested that a provision be included to allow EPA personnel to designate the appropriate approved method to be used before an applicant performs the comparison tests. One region not only did not want a reference method, but they did not want to show comparability to an approved method. It was stated that often the cost of purchasing the instrumentation to run an approved analysis obviates the need for applying for the variance in the first place. This opinion was echoed by one of the state agencies. Another state commenter felt strongly that no reference methods be designated but rather that specific precision and accuracy standards be established for given concentration ranges. It was recommended that if reference and equivalent methods are specified, they be reviewed yearly and new methods added if appropriate.

Most of the commenters agreed that data on the accuracy of the proposed alternate method should be required. One commenter suggested that variances be considered for methods that are more accurate than (but not comparable to) the approved method. Others suggested that accuracy or bias measurement would be useful in evaluating interference effects and helpful in the process of deciding whether to grant a variance. One commenter suggested that a more elaborate study of interference effects also be included in the equivalency test procedure.

Six of the regions felt that a single protocol for NPDES and NIPDWR monitoring would greatly simplify their work. One QA Coordinator suggested we pattern our protocol after the EPA Ambient Air Quality program.

One of the commenters felt that the EPA Interim Protocol placed too much emphasis on statistics and not enough on the technical aspects of the proposed alternate procedure. He suggested that a "peer review" action be included in the equivalencv evaluation process. Interested, qualified chemists would be invited to review the proposed procedures and submit their comments to the evaluating group for their consideration. Several coordinators agreed with this idea.

One suggestion was that the peer review committee be made up of EPA, state, industry and university technical personnel. The Director, EMSL-Cincinnati would act as the review judge. All criticisms of the candidates method would have to be backed by data and/or experience and the applicant would be required to respond to the critics. It was felt this system would utilize the expertise of environmental analysts and allow for more technological review of alternate methods.

Another suggestion was to establish a peer review similar to that used by ASTM. In this instance, it was felt that this peer review should be part of the proposed method development program and be done at the applicant's expense.

Sampling Cost--
All of the regions and states agreed that obtaining samples to meet the interim NPDES Alternate Method Protocol for nation-wide application was expensive and, in some cases, difficult because of potential legal problems concerning the reporting of data to an EPA agency about a monitored outfall. It was suggested that samples no longer needed by regional laboratories be held and sent, upon request, to the applicant. These samples represent a wide variety of industrial effluent categories, are relatively stable (because of preservatives or age), and their history is known. Use of these samples would eliminate the problem of gaining access to the wastewater from several industries (a significant problem for non-governmental laboratories) and reduce the cost of sample acquisition. If the age of the sample exceeded recommended holding times for the specified parameter, spiking with standards could be used. Even "old" samples would allow determination of the wastewater matrix effect on the specified parameter if spiking with standards could be used. This system would greatly reduce the cost of applying for a nationwide variance. Another regional laboratory thought that this was a reasonable solution to the problem of sample acquisition as long as the name of the industry from which the sample was taken was not disclosed. Two QA Coordinators indicated that they, however, would rather have a centralized sample bank for use by nationwide variance applicants since they were short of storage space.

One person indicated that he did not want to provide this service at government expense. Another commenter felt that the analytical equipment manufacturers should obtain samples at their expense and that a cooperative approach by industrial associations would be a better sample source. It was also suggested that testing be performed by the state agency or EPA to reduce the overall expense to the applicant.

Alternate Method - Scope of Application--
The scope of use for an approved alternate procedure was a major source of concern. Most of the regions felt that if a method were approved for use in one region, it should be approved for use in all regions. However, three regional QA Coordinators wanted the option of disapproving methods for their region if the method could be shown to be technically inadequate.

Two of the laboratories suggested that the number of industrial effluent categories samples be increased to better test matrix and interference effects. They suggested a reduction in the number of replicates to keep the number of analyses at a reasonable level. One of the states suggested that permit holders who perform a limited number of analyses in a year be allowed to submit data from a reduced number of analyses - say 5 or 10 - when applying for a variance.

SECTION 5

REVIEW OF STATISTICAL METHODS

DEFINITION OF BASIC STATISTICAL REQUIREMENTS

The brief review of the state-of-the-art in statistical evaluation aided in the identification of various components necessary to establish equivalency between two sets of data. These components provide a data evaluation system the flexibility needed to handle a wide range of data, as well as a conservative estimate of method comparability. The following statistical requirements were identified as those which may be necessary:

1) ability to determine sample number.
2) ability to identify and reject obvious errors in the data.
3) ability to determine if the data is normally distributed.
4) ability to transform data to a normal distribution when needed.
5) ability to utilize non-parametric tests.
6) ability to test for equality of means and variances.

A preliminary chart of the interrelationships of these procedures is shown in Figure 1.

Throughout the evaluation, various statistical methods which could be used to meet these requirements were identified. The review of these techniques was performed with emphasis on the advantages and disadvantages of the alternative approaches. Important consideration was given to the power, flexibility, limiting assumptions, ease of application, and the number of samples required for use.

IDENTIFICATION OF STATISTICAL COMPONENTS

Number of Samples

The choice for the number of samples must be based on a compromise between technical, economic, and statistical considerations. The technical argument is that a nationwide variance should evaluate as many different sample sources as possible to properly assess the effects of matrix interferances. The economic argument is that large numbers of samples or replicates are expensive to obtain and analyze. The statistical argument is that a large amount of replication improves the estimates of precision and accuracy. Compromise is necessary for a reasonable and economical number of analyses to be specified. It may ultimately be necessary to sacrifice statistical power for a larger variety of samples, consistent with a "real world" environment.

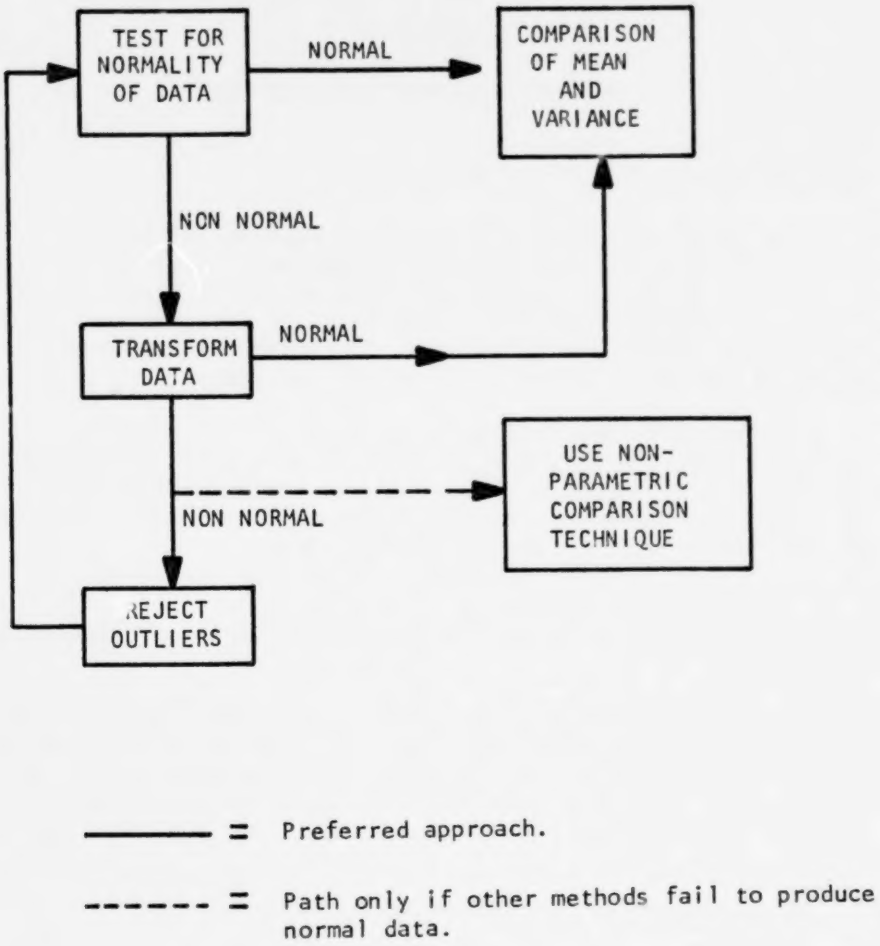

TEST FOR NORMALITY OF DATA
NORMAL
COMPARISON OF MEAN AND VARIANCE
NON NORMAL
TRANSFORM DATA
NORMAL
USE NON-PARAMETRIC COMPARISON TECHNIQUE
NON NORMAL
REJECT OUTLIERS
= Preferred approach.
= Path only if other methods fail to produce normal data.

Figure 1. Preliminary data handling approach.

Outlier Rejection

Four methods for systematic outlier rejection were reviewed: the t-test; Dixon's Test; the Range-Standard Deviation Ratio; and the Sample Moments Test. The t-test is commonly used in EPA and ASTM interlaboratory studies. The Sample Moments Test is traditionally used to evaluate normality, however, it can also be used as a powerful tool for iterative outlier rejection. The advantages and disadvantages of each method are listed in Table 4. Overall, the Sample Moments Test appears to be the most powerful test for a wide range of data and has the distinct advantage of testing for normality simultaneously with outlier rejection. The more common t-test has widespread use but has some disadvantages due to its limiting assumptions.

Dixon's test is not particularly powerful, and the Range-Standard Deviation Ratio requires evenly displaced extremes not common in wastewater chemistry.

TABLE 4. ADVANTAGES AND DISADVANTAGES OF OUTLIER REJECTION TESTS

Method	Advantages	Disadvantages
t-test	1. Widespread use in environmental analysis.	1. Assumes a normal distribution and additive error effect.
	2. Simple test.	2. Loses power as the precision term in the denominator increases.
	3. Evaluates both ends of the distribution.	
Dixon's Test	1. Fast, simple test.	1. Ranking of the data required.
		2. Not powerful. (57)
Range - Standard Deviation Ratio	1. Powerful for evaluating distributions with evenly displaced extremes.	1. Requires as much time as the t-test.
	2. Most sensitive to shewed data.	2. Evenly displaced extremes are not normally present.
	3. Simple test.	
Sample Moments Test.	1. Powerful test (57)	1. Complex calculations.
	2. Evaluates normality and rejects outliers.	2. Lack of general use.

Testing for Normality

Most statistical methods for comparing data are based on the assumption that data are normally distributed. Therefore, the ability to determine if the data meets this assumption is important. If the underlying distribution is non-normal, the proposed statistics are reduced in power. Three types of tests for normality are:

1. Empirical tests.
2. Goodness of fit tests.
3. Moment tests.

An example of an empirical test is the plotting of cumulative frequency of the data on probability paper. A straight line indicates a normal distribution. Goodness of fit tests compare the actual distribution of the data to a theoretical distribution. Shapiro, et.al., (58) state that goodness of fit tests have relatively poor power for skewed, long-tailed distributions. Moment tests measure the skewness and kurtosis of the data distribution (kurtosis is a measure of the sharpness of the peak). Shapiro, et.al., (58) indicate that moment tests are powerful for all distributions. A listing of the advantages and disadvantages of the various methods are presented in Table 5.

Each of the methods have important advantages and disadvantages which affect their application. The most critical factors are the power of the test, its limitations with regard to underlying distrubutions, the size of the population, and the difficulty of data manipulation. Using these features as governing criteria it can be seen that the cumulative frequency plot and normalized histogram do not exhibit much power and are actually preliminary tests. The common Chi-Square test requires a large number of samples and is not very powerful. The Kolomogorov-Smirnov (K-S) test also lacks power and is difficult to use. All moment test have the disadvantage of not being commonly applied. The "A" Test requires that data be placed in ranges which requires additional time for numerical manipulation. The Sample Moments Test is generally weak against certain distributions but is easier to calculate. The associated limitations with respect to different data distributions is not considered critical since much NPDES data has been found to be log normally distributed (62). The Sample Moments test exhibits good power for smaller populations and can also be used to reject outliers.

Transformations

The purpose of data transformation is to change the scale of the data in order to make the distribution normal. Most comparative tests requre the assumptions of normality, independence, and equal variance of the data. Non-normal data can be either transformed to normal data or analyzed with non-parametric tests. In an ideal transformation (63):

(a) The variance of the transformed data should be independent of changes in the mean level.

TABLE 5. ADVANTAGES AND DISADVANTAGES OF NORMALITY TESTS

Method	Advantages	Disadvantages
Cumulative Frequency Plot	1. Easily performed.	1. Preliminary or rough test. (59) 2. Visual evaluation. 3. No guidelines available to determine significance of departure.
Normalized Histogram	1. Easily performed. 2. Common test.	1. Preliminary or rough test. (59) 2. Ranking of data required. (59)
Chi-Square Test	1. Common test. (60)	1. Only approximate, especially for populations less than 50. (60) 2. Involved calculations to create the theoretical distribution. 3. Not powerful. (58)
Kolmogorov-Smirnov (K-S)	1. Any population size can be used. (61) 2. Some power for discrete distribution. 3. More powerful than Chi-Square Test. (61)	1. Relatively poor power overall. (58) 2. Involved calculations to create the theoretical distribution.
(A) Test	1. Relatively powerful test. (59) 2. Measures kurtosis.	1. Data must be placed in ranges. 2. Involved calculations. 3. Not accurate for populations less than 10. (59)
Sample Moments Test	1. Powerful test. (57) 2. Measurement of both skewness and kurtosis adds to power of test. 3. Relatively simple calculations. 4. Populations as small as 5 can be used. (57) 5. Can also reject outliers.	1. Coefficient of skewness exhibits low power against symmetric finite range distributions and very poor power against discrete distributions (57). 2. Coefficient of kurtosis exhibits low power against skewed and discrete distribution. (57)

(b) The transformed data should be normally distributed.

(c) The transformed scale should be one for which an arithmetic average is an efficient estimate of the true mean level for any particular group of measurements.

(d) The transformed scale should be one for which real effects are linear and additive.

Transformations can be performed before rejection of outliers since transformation may be all that is required to induce a normal distribution without losing data.

Basically, the empirical, or trial and error, method involves plotting data on probability paper and establishing a best fit line by trial and error. The resultant line equation provides the required transform. It is a simple method but can be time consuming. Another technique is to utilize background knowledge of the probable underlying distribution. This method is simple and fast but requires prior knowledge of the data distribution. The final approach involves an evaluation of the mean-variance relationships among the data points (64). If these relationships can be defined, and hold for NPDES data, this method could be the most effective. However, it must be remembered that transformations do not always deal effectively with the problem of non-normality in only one of the data sets (for example if data from one technique is established as normal and data from another is not). Non-parametric tests may be needed for this situation.

Non-Parametric Tests

Non-parametric tests are applicable in situations where transformation does not result in normalized data. Most non-parametric tests are not really distribution-free, since the tests are usually best suited for one type of underlying distribution and poorly suited for other types (65). These tests exhibit relatively poor power, especially when the underlying distribution is different from that assumed (65). Most require the ranking of data prior to comparison. The advantages and disadvantages of the various methods are listed in Table 6.

Other non-parametric methods have been developed, but they also lack power or have other restrictions. Since many nonparametric techniques are dependent on the underlying distribution, the knowledge of the assumed distribution is critical. It has been indicated that most NPDES parameters are log normally distributed (62). The Walsh Test appears to be the most applicable, since it has fewer underlying assumptions and more power than other procedures.

Method Comparison Test

The testing of equality of methods can be divided into two categories: comparison of means and comparison of variances. The Student's t-test and its variations and Youden's Technique compare the means directly. The use of the Student's t-test requires assumptions of normal distributions, independence, and equal variance. This t-test is a standard method for comparing means and is in common use. The calculations are simple and easily performed. A variation of the t-test, the paired-sample t-test, overcomes some of

TABLE 6. ADVANTAGES AND DISADVANTAGES OF NON-PARAMETRIC STATISTICAL TECHNIQUES

Method	Advantages	Disadvantages
Wilcoxon Test	1. Easy calculations.	1. Not powerful. (65)
	2. Best for log distributions. (65)	
Median Test	1. More powerful than Wilcoxon Test. (65)	1. Not powerful, especially for normal distribution. (65)
	2. Easy calculations	
	3. Best for double exponential distributions. (65)	
Rank-sum Test	1. Substitute for two-sample t-test. (66)	1. Assumptions required for t-test must be satisfied. (66)
	2. Not overly sensitive to unequal variances. (66)	2. Limiting efficiency of 95.5 percent. (66)
Signed Rank Test	1. Data does not have to be ranked. (61)	1. Not powerful. (61)
Walsh Test	1. Fewer limiting assumptions; only symmetry required (67).	1. Not as powerful as paired sample t-test, even when population is non-normal (59).
	2. Current usage as non-parametric test by EMSL.	2. Determination of test statistic is complex (68)
	3. More powerful than Signed Rank Test. (67,68)	3. Occurrence of symmetric, non-normal populations may not be common in wastewater analyses.
	4. Population sizes as small as four can be used. (67,68)	4. Special extension of Signed Rank Test only. (68)

the disadvantages of the t-test since the assumptions of equal variance and independence of the errors are not required (66). The test compares the differences of paired observations from two sets of data. A disadvantage of the paired-sample t-test is a loss of sensitivity due to the lower degrees of freedom for the estimates (66).

Youden's Technique is similar to a paired-sample t-test. Youden's Technique uses a two-sample chart to separate random errors, inherent systematic errors in the procedures, and systematic errors inherent to the individual laboratories. Youden's Technique has several advantages (69): a minimum number of replicates is required; more data does not automatically increase the information obtained; and calculations are simple and easily computerized. The technique has several disadvantages, especially for this application. The assumptions of normality, independence, and equal variance of the errors are required (69) and a number of collaborators or repeated testing is necessary to use the two-sample chart.

Analysis of variance (ANOVA) compares the variation due to changes in the variables tested to the variation due to random error. An F-test on the ratio of errors determines the significance of the difference and is used to test the hypothesis of data equality. The use of blocking allows data evaluations which define effects caused by concentration (i.e. samples) or analytical methods. The ANOVA has more degrees of freedom available to it for a given data set than either the regression or the paired t-test. This allows the ANOVA to set more exacting test intervals. A disadvantage of the ANOVA is that missing data complicates the design implementation.

A regression fits a least squares line to the data and measures the goodness of fit of that line to the observed values. Critical tests of hypothesis may be applied to the predicted dependent variable response, the correlation coefficient (r), the intercept and the slope.

Hilton (70) gives five critical assumptions necessary for the successful application of the regression model. The discussion of these assumptions and the consequences of violating them are discussed in terms of the deviation (errors) about the least squares fitted line.

1) Normality: The distribution of the error terms about the regression line should be normal if accurate inferences are to be drawn from the confidence intervals applied to the regression model.

2) Zero mean: The mean of the distribution of each of the error terms is zero, i.e. the mean of the error distribution falls on the regression line. Taken in concert with the first assumption this means the error terms are located symmetrically about the regression line.

3) Homoscedasticity: All of the variances of the error terms are constant. The factors causing the error terms cannot be related to the independent variable if the test statistics are to be accurately interpreted.

4) No autocorrelation of the error terms: No error term should be related to its preceding or secceeding error term. The Durbin-Watson d-statistic is an effective way to test for serial correla-

tion in the error terms. A d-value can range from 0 to 4. In the absence of serial correlation the value is 2. (71) The calculated d-value may be tested against critical values to determine if the extent of autocorrelation found in the model is significant. The correlation of error terms is defined as ρ (rho). The d- value is approximated from ρ by the equation: d approximates $2(1-\rho)$. The hypothesis tested by the d-value critical values is $H_o : \rho=0$, or that serial correlation does not exist. There are two critical regions for each d-value, d_L (d-lower) and d_u (d-upper). These are illustrated in figure 2 for a regression model with one independent variable, 24 observations (X, Y pairs) and at the 1% significance level. The region denoted by the "?" is a region of uncertainty. All d-values which are less than 1.04 or greater than 2.96 indicate significant autocorrelation of the residuals and the regression model should be considered suspect.

5) Nonstochastic independent variable: The regression model assumes the X or independent variable is known without error. In most applications this is not possible. A rule of thumb for dealing with error in the independent variable, according to Daniel, et.al. (72), is that the least squares regression can be used safely if the variance of the independent variable is less than a tenth of the average scatter of the independent variable observations from their mean. The least squares fit may also be tested by evaluating the data using Bartlett's grouping technique, which assumes error in the independent variable (73).

The advantage of the regression is the information it provides through the model equation. The disadvantages are the assumption of an errorless independent variable (also called a model 1 regression) and evenly distributed observations over the range of the variables.

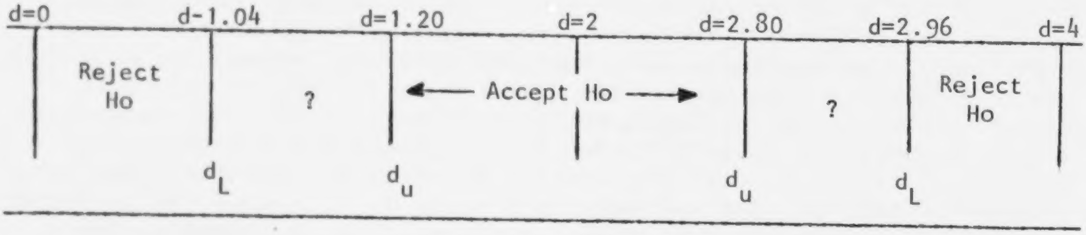


Figure 2. Critical regions of the Durbin-Watson d-value 1% significance, n=24, 1 independent variable.

Comparison of Variance

The precision of two analytical procedures can be compared by evaluating their variances. The testing of the homogeneity of variances is also needed in order to estimate a pooled variance required for the ANOVA.

Cochran's C-test is commonly used for variance testing. It is most applicable if a single variance is suspected of being appreciably greater than the rest. However, it is not as flexible as Bartlett's B-test (60), which is appropriate where random non-homogeneity is suspected between the variances. Bartlett's Test follows a Chi-Squared distribution when five or more degrees of freedom are present (60).

An F-test can also establish if a difference between two variances is significant. To compare a set of variances, a standard or reference variance must be designated and each variance individually tested. The F-test is commonly used for comparing variances. It is included in Youden's Technique, regression evaluation and ANOVA. The assumptions of independence and normal distribution of the data are required for the F-test to be more powerful(64). However, Cochran (74) has shown that the F-test is fairly robust against non-normality. The main disadvantage of the F-test is that only pairs of variances can be compared at one time as opposed to the other tests that can compare any number of variances simultaneously.

GENERAL CONSIDERATIONS

In addition to review of the various statistical methods, there are other general statistical concerns that must be addressed. Among the important subjects are the basic concepts for tests of hypothesis, confidence levels, variance impacts and the affect of not meeting the assumption of normality. These topics are briefly discussed in the following paragraphs.

Tests of Hypothesis

A statistical procedure for evaluation of data from two analytical methods to determine if they are comparable involves one or more steps in which a test of hypothesis is performed. Using statistical inference procedures, the statistician can make one of two choices: 1) rejection of the hypothesis or 2) failure to reject the hypothesis. If the hypothesis being tested was that the statistical parameters (e.g. mean) being compared were equal, rejection of the hypothesis means that the difference between the parameters was significant (at the statistical limits chosen by the statistician). Failure to reject the hypothesis means that the two statistical paramethers cannot be considered to be different (at the chosen statistical limits). However, it should be noted that failure to reject the hypothesis does not necessarily mean that the parameters are equal; it only means that a significant difference could not be found. It is possible, for example, that a real difference may exist between two parameters but, because of a high variance, the statistician may fail to reject the test hypothesis. Thus, poor analytical precision can result in the failure to detect a difference between two methods which produce dissimilar results.

Confidence Intervals

The statistical limits chosen by the statistician also influence the results of statistical inference. Tests for the homogeneity of variances or the equality of means by the t-test commonly use the 5 percent significance level (95 per cent confidence level). At this level, the statistician takes a 5 percent risk that the statistical results indicate a significant difference in the results from two methods when, in fact, the methods produce similar results (Type I error). This risk may be reduced by using a lower significance level, but this approach increases the chances of the opposite error (Type II error): failure to detect a real difference in the two methods. For the purpose of comparing results from two analytical methods, a 5 percent significance level appears to be appropriate. When testing a set of data for outliers, however, it is desirable to reduce the chance of rejecting a valid piece of data at the expense of possibly not rejecting a true outlier (minimize Type I error). Similarily, when testing a set of data for normality, the consequences of considering a slightly non-normal distribution to be normal are less serious than the consequences of making the opposite error. For outlier testing and tests for normality, therefore a one percent significance level (99 percent confidence level) appears to be the most appropriate.

Affects of Non-Normality

Non-normal distributions are not unusual in wastewater analysis. Bacteriological and radiological data usually have log normal distributions. Parameters that are normally distributed at usual levels of measurement can be non-normal when measured near the detection limit. When comparing the results from two analytical methods, normality is not a problem if both sets of data are non-normal and the same transform converts both methods to a normal form. Problems do occur, however, when one set of data is normal and the other is not. In this case, non-parametric statistics could be used. However, non-parametric statistics do not have as much power to accept or reject hypothesis as standard procedures.

Close examination of the circumstances associated with establishing comparability between NPDES methods discloses another important factor. Since both analytical methods are estimating the identical population, comparable techniques should reflect the same distribution. If the data from both methods do not have the same distribution, it can be concluded that the methods are not equivalent without further evaluation.

Effects of Variance

Establishing comparability between data sets involves evaluation and comparison of the experimentally determined means and variance. Large or extremely small variances will affect the statistical evaluation. A large variance will tend to increase the chances for establishing comparability since there is a larger range of potential overlap between methods. Therefore, a way to check the measured variance against published values for NPDES methods might be a desirable feature of a recommended protocol. On the other hand, an extremely small variance greatly reduces the chances of two techniques being judged comparable. The small variance may indicate an extremely precise technique or censoring of the data prior to its submittal. If the technique is extremely precise, accuracy testing may be needed to highlight this advantage.

SECTION 6

EVALUATION OF PROPOSED METHODS

GENERAL

Once the state-of-the-art had been reviewed, and the applicable statistical methods were identified, it was necessary to evaluate the potential techniques for use in an NPDES protocol. This involved three types of evaluation. The first was consideration of the methods' advantages and disadvantages in order to identify the most powerful and applicable techniques for further study. The second stage in evaluation was to apply data to the proposed protocols to establish the actual "workability" of the various techniques. Then cost criteria in terms of analytical testing time was included in the overall evaluation. As a result of this three stage evaluation the most feasible protocols were identified along with their advantages and disadvantages. In the next section, specific NPDES implementation considerations will be presented prior to the final recommendation in Section 7.

DEVELOPMENT AND APPLICATION OF PRELIMINARY EVALUATION CRITERIA

Approach

The previous section identified a number of statistical techniques which could be used for outlier rejection, normality testing, and mean and variance comparison. It was desirable to limit the number of techniques which would be evaluated with actural data, so an initial screening of the methods was undertaken. At first, a specific rating of the techniques was proposed. However, this approach proved to be difficult. Any specific number ranking was considered to be subjective and of limited value. Therefore, it was determined that critical variables be considered for each of the statistical procedures required. Then, the various methods were evaluated based on these considerations. The results allow elimination of the poorer statistical methods and a preliminary identification of the best techniques prior to evaluation using application of real data.

Discussion of Criteria

In the original screening process, several important considerations were highlighted. Basically these were related to the various advantages and disadvantages presented in Section 5. The following paragraphs present a brief overview of the basic evaluation criteria.

Power of the Statistic--
The most critical consideration at this point was the overall power of the statistical method being evaluated. Power in this context is defined as the ability of the procedure to maximize the chances of detecting a difference between the data at certain specified confidence intervals. Application of a more powerful statistic minimizes both Type I and II errors and allows the best comparison between analytical methods.

Applicability to Distribution Types--
Another criterion considered was the basic limitation of certain statistical techniques with respect to non-normal distributions. Although it is recognized that most of the more powerful statistical methods for comparing data assume that the random error is normally distributed about the mean, it is recognized some methods have additional limitations or weaknesses for different distributions. When the distribution is not normal, the difference between the true distribution and the normal distribution becomes part of the random error. This decreases the chances of detecting a difference between the two methods. However, some statistics have been shown to be robust when the assumption of normality is violated (68). This flexibility was considered important.

Ability to Handle Missing Data--
Some statistical comparison methods require that each group of data compared have the same number of data points. In some cases, however, data may be missing because of analytical accidents or rejection of outlier values. Comparison methods that require removal of valid data from the statistical comparison result in a decreased ability to detect a difference between two methods. Procedures which accommodate missing values are better able to detect differences.

Number of Data Points Required--
In certain instances, especially with regard to outlier rejection and normality testing, the number of data points required was an important consideration. If excessive data was needed for the statistic to be utilized, then it was not considered to be a useful technique for this application.

Ability to Evaluate Strata--
It is important for this approach to be able to evaluate the effects of different concentration levels. This is especially true in evaluating methods which may have a bias at either the low or high portions of the range. If the technique will allow relatively easy manipulation to determine strata effects, this was considered an important advantage.

General Limiting Assumptions--
Other basic limiting assumptions such as the equivalence of variance and additive error effects were also considered in the evaluation. The fewer limiting assumptions, the better.

Difficulty in Calculation and Common Usage--
These criteria were also included but their impact on the overall evaluation was limited. It was determined that calculations could be performed by

computer and that common usage was not as important as overall statistical power.

Application of Criteria

The pertinent criteria were then applied to the various statistical tests which could be performed to accomplish the needed function. Various methods were identified for use in outlier rejection, normality testing, data transformation and comparison of variance and means. The merits and weaknesses of the various statistical tests have been dicussed in Section 5.

Identification of Feasible Schemes

From the screening evaluations, various statistical methods were identified as potentially feasible for use on an NPDES protocol. These methods are identified as follows:

Outlier Rejection:	Sample Moments Test t-test
Normality Testing:	Sample Moments Test Histogram
Data Transformation:	Historical Mean/Standard Deviation Ratio
Variance Homogeneity:	Cochrans C-test Bartletts B-test
Non Parametric:	None
Method Comparison Tests:	Regression Analysis of Variance (ANOVA) Paired t-test

Basically these procedures were applied in this sequence: 1) test for normality, 2) data transformation, 3) retest for normality, 4) reject outliers, 5) retest normality and retransform, 6) comparison of variance and mean.

APPLICATION OF STATISTICAL METHODS TO DATA

Data from two NPDES alternate method applications for chemical oxygen demand (COD) were compared using the statistical tests previously described. These data provided a common point of reference for the comparison of the statistical procedures in the EPA Interim Protocol and the procedures being considered for the proposed protocol. A summary of the raw data and basic statistics utilized in these calculations is listed in Table A-2 in the Appendix. These data were evaluated as received from EPA, EMSL-Cincinnati. The effluent sources sampled by each applicant were identified by their general industrial category. It should be noted that the samples did not come from the same effluent source, although in the case of the textile category, they have the same designation.

The data was submitted by two different applicants for the purpose of obtaining a nation-wide variance to the NPDES approved reflux COD method. This statistical comparison assumed the data were submitted without prior censoring.

The term "sample" in this report meant one of six samples taken from the specified industrial category effluent. These six samples were supposed to cover the normal concentration range for COD in the effluent sampled.

The term "replicate" was applied to one of four analyses performed on each sample by each analytical method.

The data group designated COD-A resulted from the analysis of five different industrial effluent categories (Table 7). The data group designated COD-B was generated in a similar manner except that four industrial effluent categories were sampled (Table 8).

A desirable feature of any statistical method is a test of hypothesis. Each of the statistical methods evaluated, with the exception of the normalized histogram, allowed for such a test of hypothesis against the appropriate critical values given in Table 9. An asterisk (*) beside a reported value in Tables 7 and 8 indicated the value lay outside the specified critical region of that test. Statistically this would indicate rejection of the tested hypothesis. Values within the critical region indicated failure to reject the hypothesis. The discussion of the applied data, however, was conducted in terms of comparability. Methods which were judged comparable were those which failed to reject the test hypothesis. Often two or more statistical techniques failed to agree on comparability. An attempt was then made to relate these inconsistencies to the power of the test statistics or their limiting assumptions. It was not possible to determine which of two disagreeing tests were "right". An attempt was made to infer from residual plots, degrees of freedom available and related statistical tests which techniques were most likely correct when the methods yielded different results.

The interpretation of most of the statistical tests used in Tables 7 and 8 requires no explanation. The interpretation of the histogram and the regression model is not as apparent. The seven cell histogram used to evaluate normality was the most subjective statistic applied. Sample distributions were considered non-normal (NN) if they looked skewed or displayed other non-gaussian traits such as bimodality. No critical test regions were evaluated.

The regression model tested comparability using the coefficient of determination (r^2), the slope (b_1), the intercept (b_o), and the Durbin-Watson d-value. Two methods were considered comparable by the regression model if a t-test indicated the slope was equal to one, the intercept was equal to zero and if the F-test indicated the r^2 value was not zero. The regression model was considered appropriate for data comparison if serial correlation of the residuals as measured by the d-value did not exceed the d_L limit values given in Table 9.

TABLE 7. A COMPARISON OF STATISTICAL RESULTS USING COD-A DATA

Effluent Source	Method[a]	Mean	Range	Normality			Variance		
		$\overline{X}$		Hist.	$\sqrt{b_1}$	b_2	Cochran's	Bartlett's	F-Test
Mining	A	44	16-64	N	-0.05	1.15	0.329	5.300	2.0212
	P	43		N	0.05	1.21	0.479	9.200	
Petroleum	A	47	31-83	N	1.06	2.96	0.917	29.066*	2.283
	P	48		N	1.01	2.58	0.622	7.998	
Inorganic Chemical	A	1200	80-3,300	N	1.35*	3.61	0.492	16.015*	1.008
	P	1200		N	1.24*	3.65	0.778	22.295*	
	A-Log	750			0.97	2.80			
	P-Log	740			1.03	2.61			
Plastics Industry	A	460	44-1,400	NN	0.74	1.59		17.750*	
	P	460		NN	0.74	1.61		14.041	
	A-Log			N			0.819		6.868*
	P-Log			N			0.497		
Textile	A	350	270-630	N	1.11*	3.11	0.352	1.474	3.412
	P	350		N	1.00	2.94	0.482	9.200	
	A-Log	320			0.50	2.89			
	P-Log	327			0.37	2.16			
Total	A	420	16-3,300		2.82*	4.95*	0.386*	106.27*	1.135
	P	420			2.92*	4.83	0.305*	78.98*	
	A-Log				1.12*				
	P-Log				1.37*				

a) A=NPDES Approved Method.
P=Proposed Method.

*Indicates the reported value exceeds the critical test interval.

(continued)

TABLE 7. (continued)

Effluent Source	Method[a]	Methods Comparison Test							Comparability Results (P=A)		
				Regression							
		t-test t-value	ANOVA F-Value	r^2	Intercept	Slope	Auto-corr	d-value	t-test	ANOVA	Regres.
Mining	A	0.435	0.450	0.963	2.550	0.927	0.025	1.820	YES	YES	YES
	P										
Petroleum	A	0.862	0.637	0.954	-2.024	1.070	-0.229	2.455	YES	YES	YES
	P										
Inorganic Chemical	A	1.257	12.570*	0.999	-26.43*	1.048*	0.591	0.729*	YES	NO	NO
	P									(Log)	
	A-Log		0.598	0.999	-0.157*	1.053*	0.801	0.364*		YES	NO
	P-Log										
Plastics Industry	A		4.96*	0.999	9.389*	0.989*	0.235	1.523		NO	NO
	P									(Log)	
	A-Log	2.776*							NO		
	P-Log										
Textile	A	1.969	26.5*	0.998	0.683	1.021*	0.573	0.84*	YES	NO	NO
	P									(Log)	
	A-Log		1.236	0.997	-0.054	1.024	0.536	0.910*		YES	YES
	P-Log										
Total	A	1.674	20.800*	0.999	-4.195*	1.030*	0.603	0.795*	YES	NO	NO
	P									(Log)	
	A-Log			0.995	0.024	0.994	0.310	1.379*			YES
	P-Log										

a) A=NPDES Approved Method.
P=Proposed Method.

*Indicates the reported value exceeds the critical test interval.

TABLE 8. A COMPARISON OF STATISTICAL RESULTS USING COD-B DATA

Effluent Source	Method[a]	Mean	Range	Normality			Variance		
		$\overline{X}$		Hist.	$\sqrt{b_1}$	b_2	Cochran's	Bartlett's	F-Test
Textile	A	47	28-86	N	1.82*	6.36*	0.778*	18.462*	1.152
	P	56		N	0.93	2.55	0.425	15.024	
	A-Log	45			0.64	3.16			
	P-Log	49			-1.07*	2.94			
Domestic	A	89	26-105	N	0.89	2.83	0.467	6.421	7.825*
	P	90		N	1.33*	3.83	0.568	15.815*	
	A-Log	74			-0.34	2.86			
	P-Log	78			0.44	2.71			
Hospital	A	78	62-94	N	0.08	2.17	0.329	12.898	9.848*
	P	83		N	0.52	2.50	0.367	3.045	
Dye	A	132	81-220	N	1.70*	2.99	0.337	8.525	2.042
	P	119		N	1.22*	2.78	0.482	15.732	
	A-Log	126			0.39	2.16			
	P-Log	110			-0.25	2.23			
Total	A	87	26-220		1.21*	5.06*		120.02*	
	P	89			1.13*	4.19		135.16*	
	A-Log	75			-1.00*	3.12			
	P-Log	79			-0.97	3.24			

a) A=NPDES Approved Method.
P=Proposed Method.

*Indicates the reported value exceeds the critical test interval.

(continued)

TABLE 8. (continued)

Effluent Source	Method[a]	Methods Comparison Test							Comparability Results (P=A)		
				Regression							
		t-test t-value	ANOVA F-Value	r^2	Intercept	Slope	Auto-corr	d-value	t-test	ANOVA	Regres.
Textile	A	2.698*	11.500*	0.562	19.961*	0.781	0.073	1.850	NO	NO	NO
	P									(Log)	
	A-Log		14.800*	0.455	0.688*	0.633*	0.213	1.571		NO	NO
	P-Log										
Domestic	A	0.165	0.036	0.915	6.850	0.930	-0.283	2.560	YES	YES	YES
	P									(Log)	
	A-Log		1.490	0.877	0.481*	0.755*	0.095	1.809		YES	NO
	P-Log										
Hospital	A	1.142	6.049*	0.101*	30.11	0.703	0.551	0.893*	YES	NO	NO
	P										
Dye	A	1.772	12.190*	0.834	-8.134	0.965	0.260	1.47	YES	NO	YES
	P									(Log)	
	A-Log		14.178*	0.800	-0.414	1.170	0.290	1.41		NO	YES
	P-Log										
Total	A	0.80	1.340	0.788	16.145*	0.840*	0.337	1.325*	YES	YES	NO
	P									(Log)	
	A-Log		3.590	0.702	0.481*	0.755	0.430	1.141*			NO
	P-Log										

a) A=NPDES Approved Method.
P=Proposed Method.

*Indicates the reported value exceeds the critical test interval.

TABLE 9. SUMMARY OF CRITICAL REGIONS

Test	Number	Degrees of freedom	Critical value	Confidence interval
I. Outliers				
t-Test	4	3	4.541 (one tail)	99%
b_2	4		3.11 (n = 5)	99%
II. Normality				
b_1	24		1.06 (n = 25)	99%
	>60		0.72 (n = 60)	99%
b_2	24		5.00 (n = 25)	99%
	>100		4.39 (n = 100)	99%
III. Variance				
Cochran's		3, 6	0.626	99%
		3, 24	0.230	99%
		3, 30	0.191	99%
Bartlett's		5	15.09	99%
		23	41.64	99%
		29	49.59	99%
F-test		18, 18	2.22	95%
		72, 72	1.53	95%
		90, 90	1.35	95%
IV. Method comparison				
Paired t-test		5	2.105 (one tail)	95%
		25	1.708 (one tail)	95%
		5	2.571 (two tail)	95%
		25	2.060 (two tail)	95%
		1, 41	4.080	95%

(continued)

TABLE 9. (continued)

Test	Number	Degrees of freedom	Critical value	Confidence interval
ANOVA regression		1, 41	4.080	95%
r^2	24		0.179	95%
		>90	0.040	95%
t-slope		22	2.074	95%
t-intercept		>100	1.980	95%
d-value	24		d_u = 1.20, 2.80 d_L = 1.04, 2.96	99%
	>100		d_u = 1.56, 2.44 d_L = 1.52, 2.48	99%

RESULTS OF STATISTICAL TESTS

No outliers were identified by either the b_2 or the t-test statistic in any of the data tested. The results of the other statistical tests are discussed by industrial effluent category beginning with the data presented in Table 7 for the COD-A data. The results of the three method comparison tests (paired t-test, ANOVA and regression) are emphasized.

Results of Evaluation of COD-A Data

Mining, COD-A--
This data covered a reasonably wide range with the samples well distributed throughout the range. Data from both methods were normally distributed and had homogeneous variances over the concentration range tested. All three method comparison techniques indicated that the approved and proposed COD methods were comparable.

A scatter plot of the data from the regression analysis is shown in Figure 3. The slope and intercept of the best fit model are statistically equal to one and zero respectively (95% confidence interval) and the variance due to regression accounted for approximately 96% of the total regression. This is illustrated by the 95% confidence boundaries and the close fit of the model best fit line to the ideal fit line. Note that the observations are close to the best fit line and they are evenly distributed about the line at all concentrations. This distribution resulted in a low autocorrelation and a d-value within the critical region, demonstrating that the error terms were uncorrelated and additive.

Petroleum, COD-A--
The Petroleum effluent category was similar to the Mining, COD-A data except that the B-test indicated that the approved method suffered from non-homogeneity of within-sample variance. This lack of variance consistency was not large enough to effect the t-test, the ANOVA or the regression. It may account for the increased autocorrelation and decreased d-values. All three procedures indicated that the methods were comparable.

Inorganic Chemical, COD-A--
These samples covered the largest range of any of the industrial effluent categories (80 to 3,300 mg/1 COD). The data suffered from skewness which was corrected by log transformation. The B-test revealed the within-sample variance estimates were not equal for either method. This may have been due to the highly precise data reported for the group. The coefficients of variation ranged from 0.0% to 1.6% for the sample replicates, which is very unusual for the NPDES reflux COD method. For comparison, an interlaboratory estimate of COD precision, stated as the coefficient of variation, was 33.6% at 12 mg/1 and 6.9% at 270 mg/1 using samples prepared in distilled water. These results were based on Winter's interlaboratory demand study (10).

The precise measurements in this category set very close confidence intervals for the test statistics. The paired t-test, the ANOVA and the regression yielded different results. For the non-normal distribution, the methods

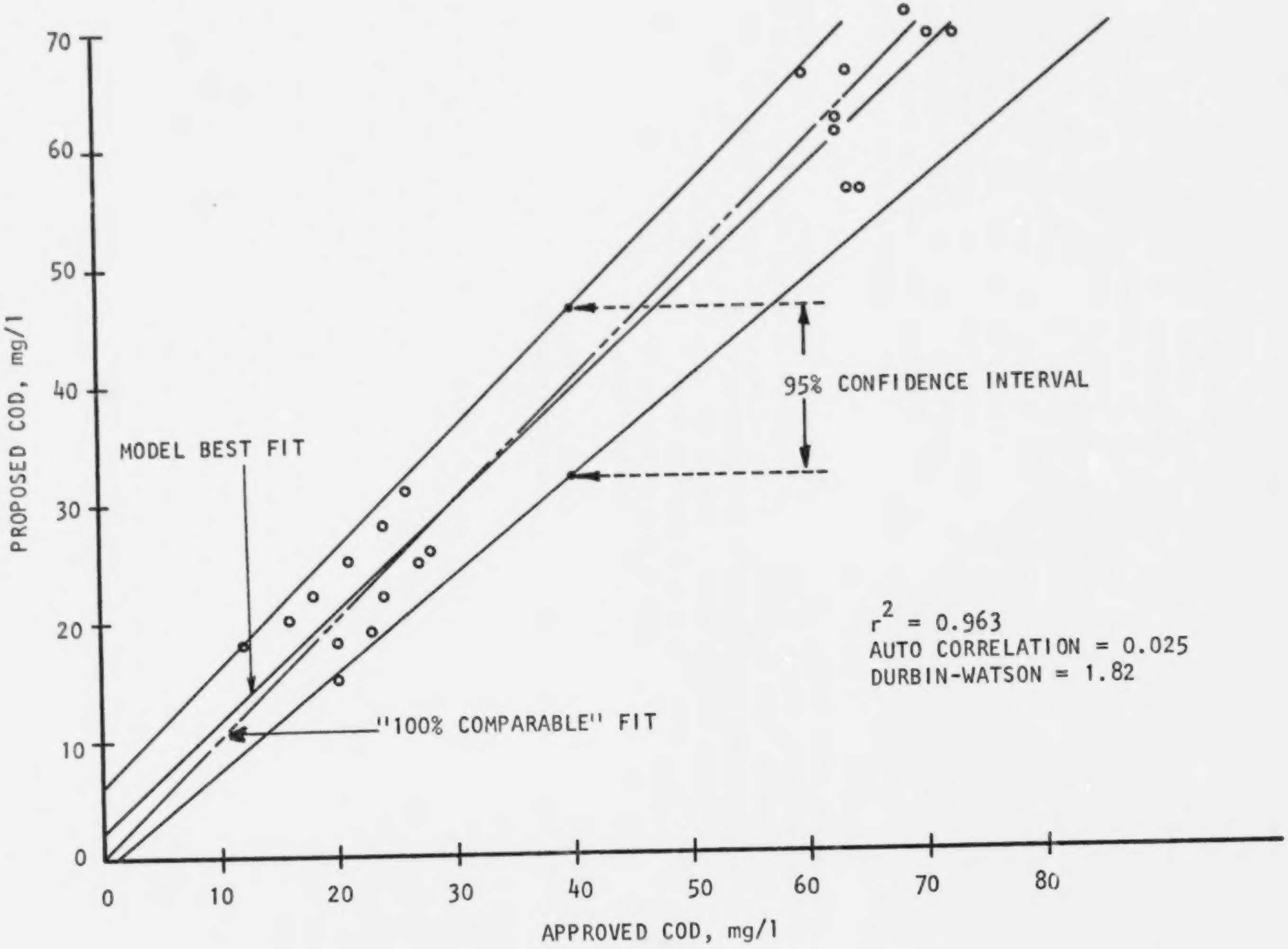


Figure 3. COD-A mine effluent scatter plot.

were judged comparable by the t-test and not comparable by the ANOVA and the regression. Although the regression was able to account for much of the variance of the dependent variable ($r^2 = 0.999$), the slope was significantly different from 1.0 and the intercept was significantly different from zero. The d-value exceeded the critical limit, which indicated serial correlation among the error terms. This serial correlation is illustrated in the residual plots (Figure 4-6). Figure 4 is a scatter plot of the COD-A inorganic chemicals data. Because the range of these data is so large, it is difficult to detect serial error correlation in the scatter plot. The data strata were expanded in Figures 5 and 6 to illustrate this effect. In Figure 6 the data in the range of 3340 to 3700 mg/1 COD and 65 to 70 mg/1 COD systematically lie above the fit line while all the mid range observations (925 to 1050 mg/1 COD) lie below the line. The regression model can fit a line which, on the average, accounts for much of the variance because it assumes error additivity and normal distribution, but the d-value cautions the analyst against placing confidence in the regression model for this data set.

After a log transformation, the ANOVA procedure indicated that the methods were comparable. The regression procedure indicated that the results from the two methods were significantly different. The autocorrelation and d-value indicated significant error non-additivity effects. The log transforms did not eliminate error interaction effects. The effect of the error non-additivity upon the t-test and the ANOVA is uncertain. The evaluation of the residual plots suggested a concentration dependent systematic error occurred in the inorganic chemical data.

More information would be necessary to determine if these biases were related to industrial source, analysts, concentration, or some other variable.

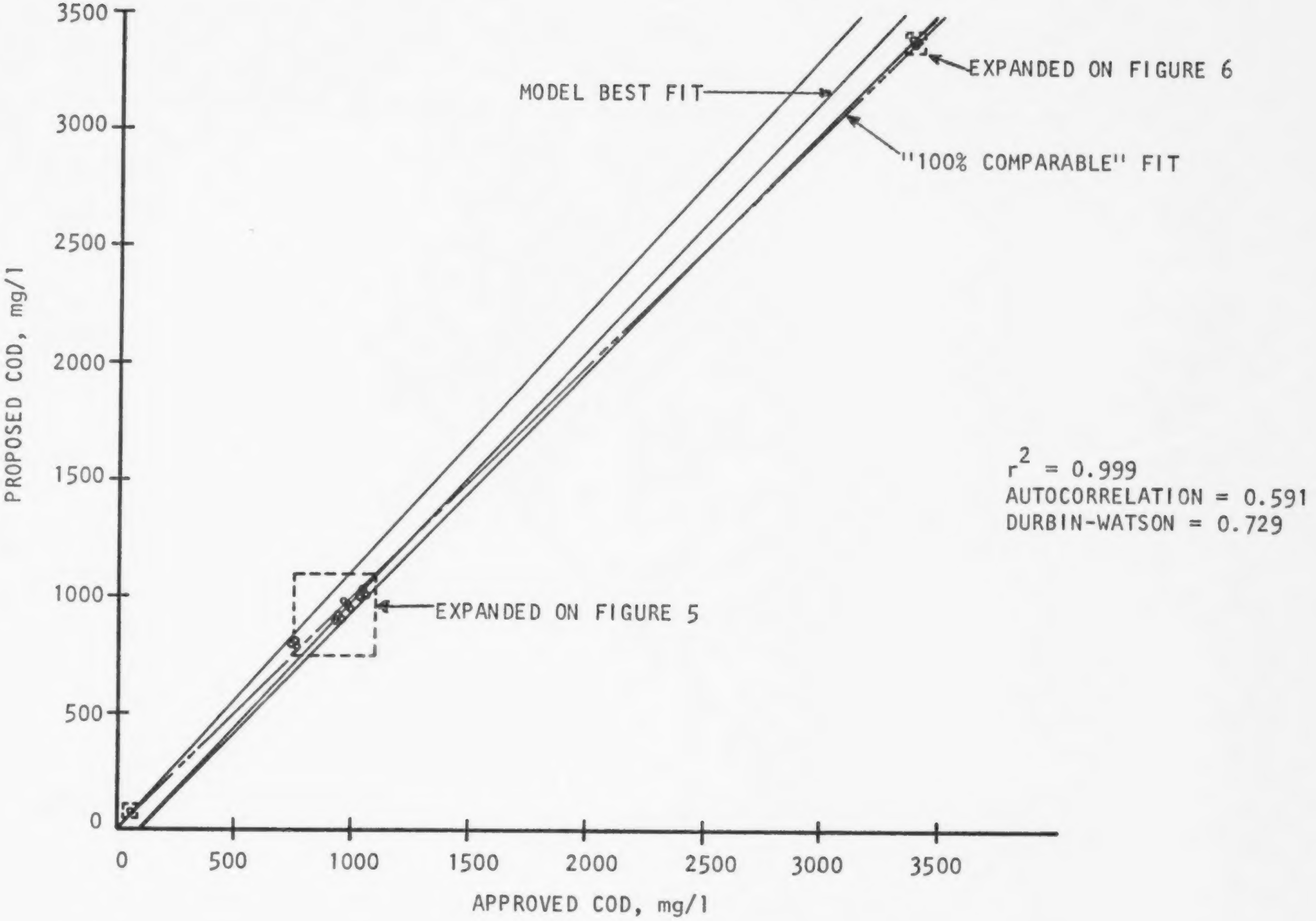
Plastics, COD-A--

The Plastics industry category samples also covered a large range (44 to 1400 mg/1 COD). The data were normally distributed according to the Moments Test but were considered non-normal by the histogram method. As a result, the data evaluation performed by the EPA Interim Protocol (i.e. the C-test, the F-test and the paired t-test) was done on the log transformed data.

The ANOVA and regression analyses were performed on the non-transformed data. The B-test showed the approved method sample variance estimates were not equal. This data, like the Inorganic Chemical data, had very high within-sample precision. Eleven of the twelve sample sets had coefficients of variation below 6%, and nine of the twelve were below 2%. Because two of the larger estimates of variance were in the approved method sample group, the F-test rejected the hypothesis of equal variances between the two methods. This trend was evident in the t-test, the ANOVA and the regression. All rejected the hypothesis that the methods were comparable.

Textile, COD-A--

The Moments Test showed the approved method data was skewed and the proposed method data was normal. After log transformation, the data was normal for both methods. Both the C-test and the B-test indicated the within-sample



$r^2 = 0.999$
AUTOCORRELATION = 0.591
DURBIN-WATSON = 0.729

Figure 4. COD-A inorganic chemicals scatter plot.

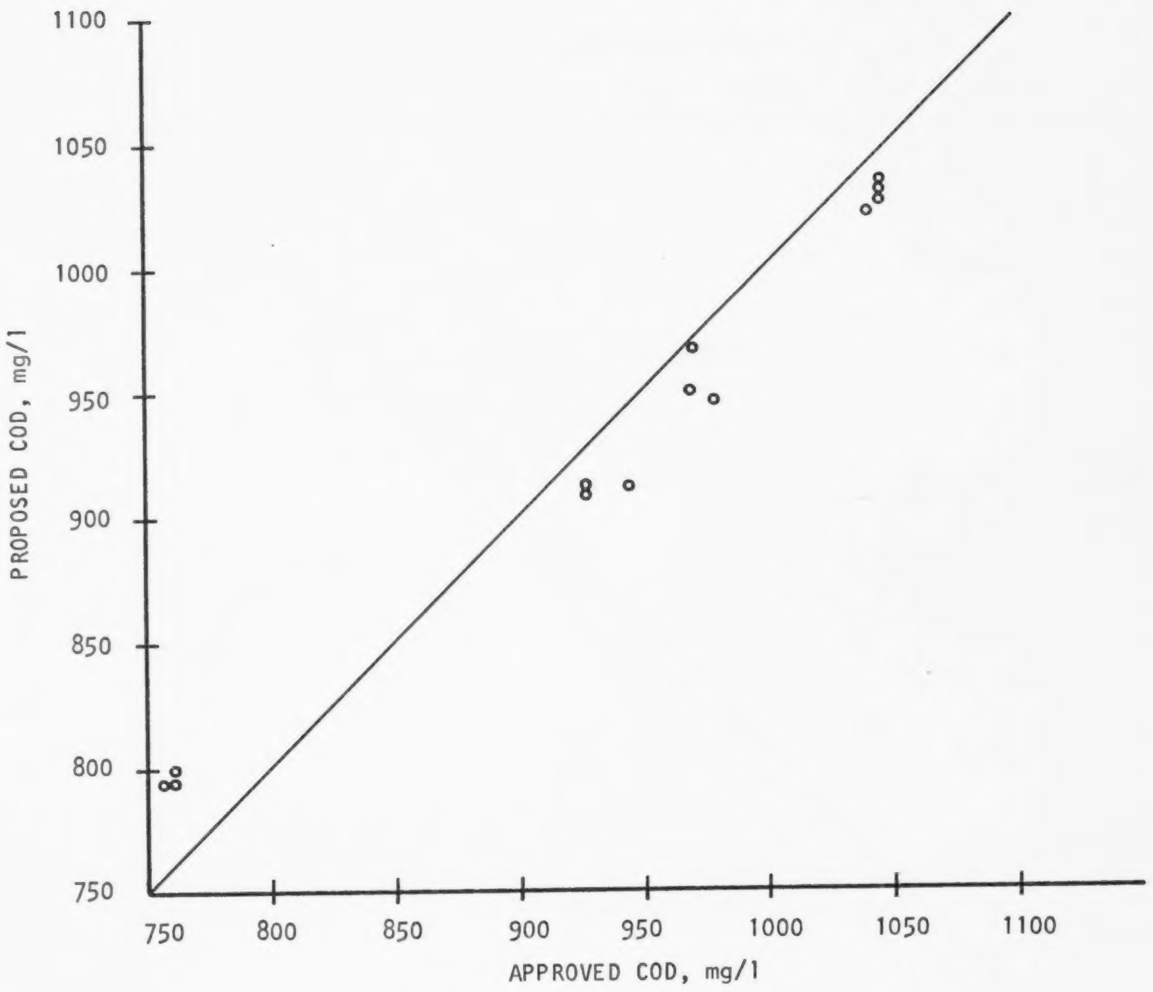


Figure 5. COD-A inorganic chemical - expanded central range.

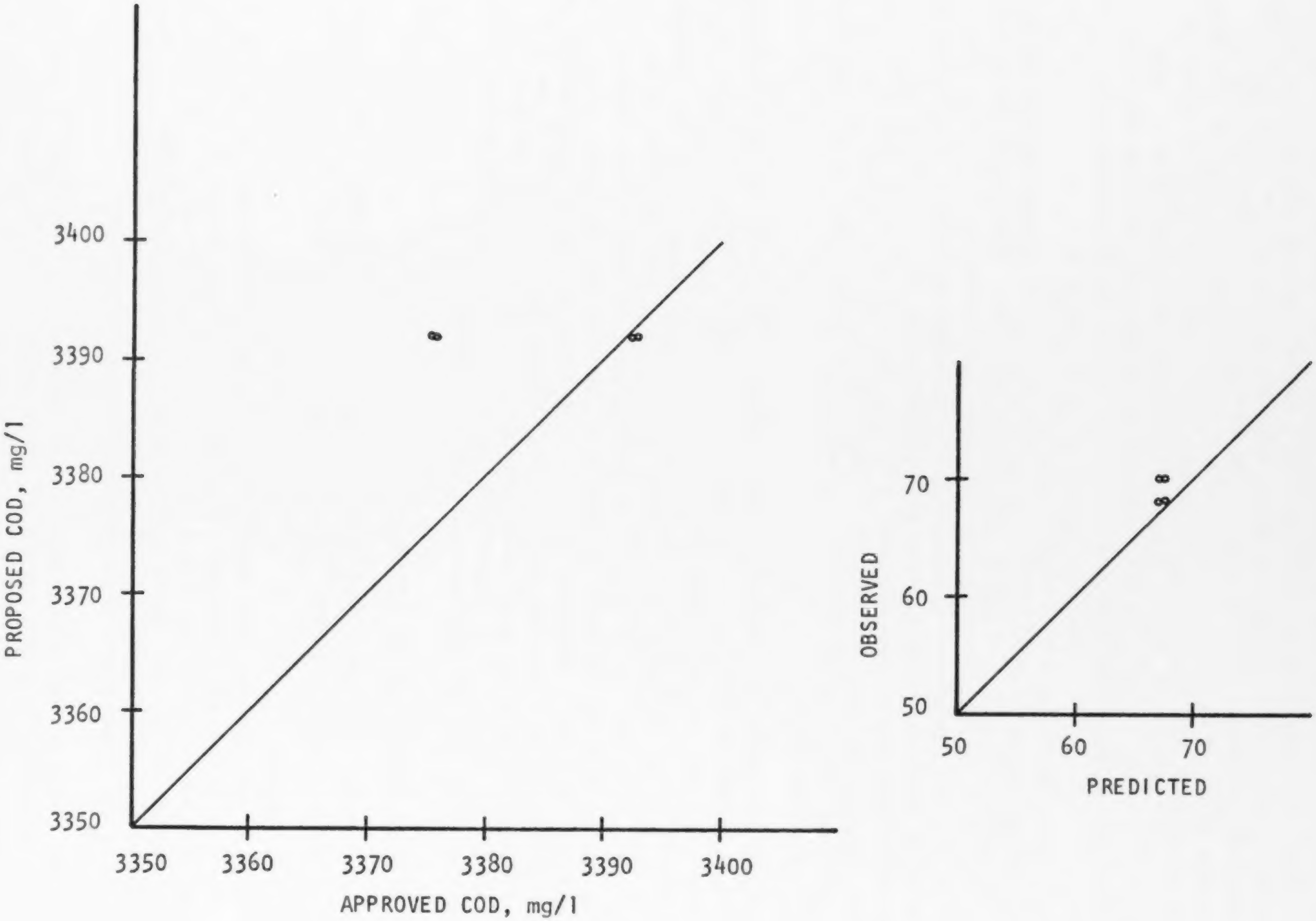


Figure 6. COD-A inorganic chemicals - expanded extreme ranges.

variance estimates could be pooled. The within-sample precision was very high. The coefficients of variation ranged from zero to 1.3%. The F-test indicated that the sample variances were significantly different. When evaluating the non-transformed data, the two analytical methods were considered to be comparable by the t-test but not by the ANOVA or the regression procedures. After transformation all methods accepted the hypothesis of method comparability. However, the d-value of 0.91 inferred a significant degree of serial correlation among the error terms. The regression model results should therefore be carefully evaluated with the understanding that an important limiting assumption has been violated.

All Categoried, COD-A--

The combined data from all the industrial effluent categories was found to be non-normal before and after log transformation. The extent to which this non-normal distribution affected the test statistics is unknown. The C-test and the B-test would not allow variance estimate pooling. The paired t-test indicated that the methods were comparable using the non-transformed data. The ANOVA and regression did not. The regression accepted the hypothesis after transformation of the data. The d-value for the non-transformed and transformed data showed there was significant error additivity about the least squares line.

The data presented by the applicant in COD-A, (Table 8) had a high degree of within-sample precision. The analytical methods were found to be comparable for four of the five industrial effluent categories when evaluated by the t-test or the ANOVA. Regression analysis showed that the methods were comparable in three of the five categories. Serial correlation among the residual error terms was significant in two of the industrial effluent categories as well as when all the data was grouped. A critical review of the data indicates that error interaction would make unambiguous statistical interpretation impossible. Historical information pertaining to the variables of time of analysis, analyst and sequence of analyses might aid in determining comparability. Data evaluating accuracy as determined from spiked samples would also be helpful.

Results of Evaluation of COD-B Data

The data submitted for COD-B was generated from four industrial categories and covered a much narrower range (26-220 mg/1 COD) than the COD-A data.

Textile, COD-B--

This data set was the only one evaluated where transformation failed to produce a normal distribution. The samples covered a limited concentration range (48 to 86 mg/1) and the within-sample precision was relatively poor for both methods. All three method comparison procedures showed that the methods were not comparable. The non-normality and lack of variance homogeneity undoubtedly weakened the statistical test.

A wider range of concentrations would allow more information in evaluating comparability of the COD method. The evidence from the given data is that the proposed method did not give results comparable to the approved method for the textile effluent source.

Domestic, COD-B--
The range of sample concentrations was nearly as limited as that of the textile category data. The proposed method data was skewed. A log transformation left both data sets normally distributed. The B-test for sample variance equality rejected the proposed data set variance estimates. The coefficient of variation ranged from 3% to 19%. This poor precision apparently accounted for contradictory results of the three test procedures. When applied to the non-transformed data, all three method comparison tests agreed the methods were comparable. After normalization of the data by transformation to the log form, the t-test and ANOVA accepted the comparable hypothesis; the regression rejected it. The reported d-value would eliminate non-additive error effects as a cause for rejection. An examination of the regression line shows that the regression model fit it's line to data that was tightly clustered in terms of concentration. The line fit the limited data relatively well ($r^2 = 0.877$), but the imposed comparability requirements of a slope of one and an intercept of zero were not met. The ANOVA and the t-test were not effected by tightly grouped data.

Hospital, COD-B--
This data is an excellent example of comparing results from a precise method (approved method data) to an imprecise method (proposed method data). These samples concentrations covered the narrowest range of all the categories reviewed (62-94 mg/1 COD). The data for both methods were normally distributed. The B-test and the C-test allowed within sample variance estimate pooling. The average coefficient of variation for the approved method was 2.3%; for the proposed-method it was 16%. The F-test showed that the difference in variance between the two methods was significant.

The paired t-test indicated that the two methods were comparable. This was probably due to the imprecision of the proposed method data and the limited number of degrees of freedom for the test. Both the ANOVA and regression showed that the methods were not comparable. The model best fit line could account for only 10% of the total variance ($r^2 = 0.101$), which was statistically equal to zero. This effect is illustrated in Figure 7 by the widely scattered data points and the poor fit of the model line and the ideal fit line. Evaluation of the slope and intercept are meaningless because the r^2 value had already determined no correlation exists between the methods.

Dye, COD-B--
Samples obtained from the Dye industrial category covered the largest range of concentrations in the COD-B application (26-220 mg/1 COD). The data for both methods were skewed. After log-transformation, both sets of data were normal. The B-test and the C-tests allowed variance pooling and the F-test accepted the pooled variance estimates as equal. Both methods had good precision with the coefficients of variation averaging 7%.

The paired t-test and the regression indicated that there was no significant difference between the results from the two methods. The ANOVA rejected the methods as comparable. An examination of the log-normal data by the regression model indicates that the fit line can account for about 80% of the total variance and the non-additive error effects are minimal. There is no crossover effect within the concentration range evaluated. The proposed

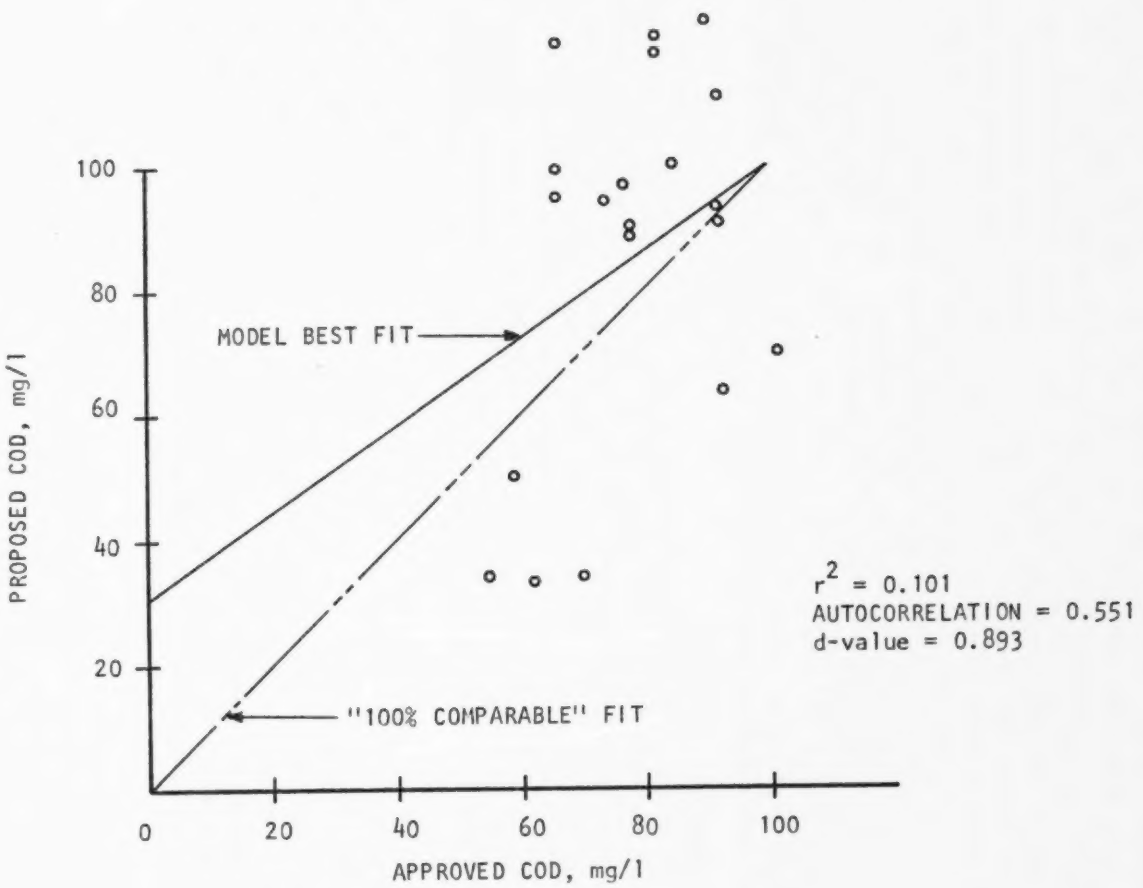


Figure 7. COD-B hospital scatter plot.

method underestimated the approved method. Crossover doesn't occur until 300 mg/l COD. The ANOVA interpreted the apparent systematic bias of one of the methods as significant. The greater sensitivity of the ANOVA in this case may be attributed to the fact the ANOVA has twice as many degrees of freedom as either the t-test or the regression.

All Categories, COD-B--
When all industrial categories were evaluated together, they could not be made normal by log transformation nor were the within-sample variance estimates equal. The consequence of failing to meet these two assumptions may have confounded the test result. The paired t-test and the ANOVA accepted the two methods as equal; the regression rejected them. The t-test and ANOVA may have been weakened by the imprecise data from the Textile, Domestic and Hospital categories. Error additively among the residuals also weakened the statistical tests. The effect of violating so many limiting assumptions is that the normal probabilities upon which the statistical tests are based would not hold for these data.

Summary

The application of the data from these two NPDES alternate method studies provided useful if not unambiguous information about the statistical tests for normality, variance equality and method comparability.

The Moments Test was more sensitive in detecting data skewness and kurtosis than the normalized histogram. Log transforms corrected skewness in all industrial effluent categories where it occured. It induced skewness in the proposed data set of Textile, COD-B. The mean/variance ratio test was not able to reliably predict an appropriate transform for these data and was abandoned. The B-test was more sensitive than the C-test for detecting heterogeneous within-sample variance estimates.

The paired t-test most often accepted the two methods as comparable. It seemed most susceptable to accepting data as comparable when one or both methods were decidedly imprecise. The ANOVA and the regression procedures generally gave similar results. The regression was more affected by data that did not cover a wide range of concentrations, or that systematically varied above and below the best fit line. The assumption of an errorless independent variable did not weaken the regression model when the independent variable was subject to considerable error (Textile, COD-B).

The ANOVA procedure was the most sensitive to bias when it occurred in one method consistently, the data variance was homogeneous and the data was normally distributed. The ANOVA had twice as many degrees of freedom as either the t-test or the regression which contributed to it's greater sensitivity.

The regression model provided the most information about how the data sets compared and which factors were important when the data sets were different. The d-value was most useful in identifying serial correlations among the error terms.

The data evaluated in these two NPDES applications often violated one or more of the limiting assumptions of the three statistical tests used to determine comparability. This resulted in conflicting conclusions from the tests of hypothesis. In most cases, it was the examination of the underlying causes of conflict which provided insights into how or why the methods differed from each other.

ECONOMIC CONSIDERATIONS

The costs of performing the work required by the equivalency testing protocol are of great importance to a laboratory when deciding whether or not to apply for a method variance. The equvalency testing protocol imposes two types of costs on the applicant; 1) the cost of performing the required number of analyses and 2) the cost of obtaining the required number of samples. Although these costs cannot be accurately predicted without details of the specific application, their impact can be approximated. This information should be considered when recommending a final protocol.

Analytical Costs

The EPA Interim Protocol specifies that six samples be obtained from each outfall tested and that each sample be analyzed four times by each analytical method. This results in a total of 48 analyses per outfall. Because of the requirement for pairing data for the paired t-test and the regression test, and the difficulties in forming pairs of the individual analytical observations, it is customary to pair the means of the sample replicates. Thus the degrees of freedom available for these method comparison techniques is dependent on the number of samples analyzed and not upon the number of replicates. Increasing the number of samples analyzed and decreasing the number of replicates is a way of increasing the power of the paired t-test and regression without increasing the total number of analyses. This approach, unfortunately, decreases the power of the ANOVA, but the loss is small. This effect is illustrated in Table 10. This approach is desirable from the chemist's point of view because 1) more samples are analyzed, increasing the ability to include more concentrations, 2) the use of two replicate analyses rather than four is closer to the manner in which the analytical methods are routinely used, and 3) the total number of analyses remains at a reasonable level.

Increasing the number of samples analyzed does not necessarily require that more samples must be taken. In fact, it would be advantageous to keep the number of samples obtained from the field at a reasonable number, such as six, and develop the additional samples by diluting, blending or spiking selected samples. This practice would keep the sampling effort at a reasonable level and would allow the analyst to prepare additional samples that cover the entire range of concentration to be tested without increasing sampling costs.

TABLE 10. EFFECT OF NUMBERS OF SAMPLES AND REPLICATES ON METHOD COMPARISON STATISTICS

Number of Methods Compared	2	2
Number of Samples	6	12
Number of Replicates	4	2
Total Number of Analyses	48	48
Paired t-test		
Degrees of freedom	5	11
Critical t-statistic (95%)	2.57	2.20
Regression		
Degrees of freedom for regression	5	11
ANOVA		
Degrees of freedom for method	1	1
error estimate	41	35
Critical F Value (95%)	4.08	4.12

SECTION 7

IMPLEMENTATION CONSIDERATIONS

The importance of developing a usable protocol for analytical method comparison cannot be overemphasized. If a complex, difficult to use, or confusing procedure is developed, it will not be readily applied. The critical nature of this problem was particularly apparent when the state and regional QA coordinators discussed their recommendations for development of an acceptable protocol. It was clearly stated that prospective applicants did not request a variance because they felt the process was too complex and difficult for them to implement. Some of the comments are addressed in the applicants user's packet which has been prepared in conjunction with this project effort, however, other problems are clearly associated with the implementaiton of the NPDES Protocol and are outside of the basic work scope of this project effort. These are not statistical in nature and are not simply solved. However, in order to provide a more viable protocol, they should be addressed. This section presents a discussion of some specific problem areas, and proposes some potential solutions. It must be understood that the proposed solutions reflect the opinions of the authors. The final solution of these implementation considerations will be determined by EPA officials when the protocol is completely developed.

REFERENCE METHOD

When method comparisons are performed, results from one proposed technique are compared to results from an approved procedure. However, in the case of NPDES approved procedures, more than one method can be considered acceptable. Therefore, it is conceivable that a proposed procedure may be comparable to certain approved techniques but not others.

One approach to solving this problem would be similar to that utilized by EPA in approving air quality measurement methods. In this protocol. specified methods and instruments are identified as reference techniques and all alternate procedures must be comparable to them. These reference techniques have undergone extensive preliminary testing prior to acceptance as the basis for all future comparisons. In order to eliminate potential problems associated with NPDES method comparison, it was considered desirable for the water quality approach to be similar. However, further evaluation indicated that the potential interferences to one method caused by a wide variety of sample matrices could be excessive. The individual analyst needed flexibility in choice of analytical procedure in order to obtain the best results for a specific sample. Specification of a single reference method, although it would simplify the list of approved methods and the comparability protocol, could actually reduce the accuracy of constituent analyses. Therefore, it was determined that a reference method could not be specified. Instead the applicant should be allowed to choose an approved technique to use in a specific comparison, as is currently being done.

ACCURACY TESTING

Although it is understood that the proposed protocol is limited to establishing comparability between analytical methods, a provision for some type of accuracy testing would allow a more meaningful evaluation of the test results. This is especially true when the proposed technique may be a more accurate measure of the parameter than the approved procedure. The statistical evaluation may indicate that the methods are non-comparable even if a new method is more accurate. In this instance a better method cannot be identified without some provision for accuracy testing.

Method accuracy is best established by the analysis of a sample containing a known concentration of a specific constituent. The EPA has developed samples of this type for many parameters in conjunction with their quality assurance program. If these samples could be incorporated into the equivalency testing program, especially when nationwide applications are involved, the accuracy for both methods could be established.

Other methods of accuracy evaluation involve spiking and dilution of samples with initially known concentrations. Although the matrix effects may not be identical to those occurring naturally, dilution is a viable technique for both accuracy determinations and sample analyses when the parameter concentration exceeds the analytical method range. Spiking is commonly used to evaluate accuracy in terms of percent recovery. As with dilution, the matrix effects may not be identical to those found in the environment, however, since both analytical methods will be performed on samples treated the same way, the relative accuracy can still be determined. A major advantage of using spiking and dilution for comparison studies is the ability to increase the range of concentrations being analyzed. Natural samples collected specifically for method comparison or other reasons may not contain constituents in concentrations which cover the range of the analytical method. Including provisions for spiking and dilution, as well as mixing and blending will allow the concentration range to be more completely covered. In this way, a better comparison of methods will be obtained and the relative accuracy can be tested.

LIMITED VARIANCES

Analysis of constituents within an industrial wastewater is extremely difficult due to the large variety of potential interferences which may be encountered in different samples. Therefore, it is not surprising that analytical methods may be comparable for some industrial effluent category discharges and not for others. Examples of this effect were shown in Section 6. A method may be excellent for all industrial discharge categories analyzed except one, and yet not be a generally comparable technique or it may be shown to be generally comparable but very inaccurate for one type of discharge. In these cases, the limitations of the technique should be known in order to allow the most effective data generation. It may then be possible to limit the variance to industries where the method was shown to be acceptable. This approach will be difficult to implement by regulatory personnel; however, it would allow more appropriate utilization of new analytical methods.

TYPES OF SAMPLES REQUIRED FOR NPDES METHOD COMPARISON

It is apparent that it is unreasonable to require sample analyses of all the discharges which may contain the parameter of interest. Therefore, it is necessary to limit the number of industrial discharge categories studied. In the development of the EPA Interim Protocol, a specific rationale was presented with respect to the choice of five industrial effluent categories in the protocol. It was determined that the discharges with the most impact on the overall NPDES program could be represented by five industrial categories. The following information was developed (72).

> "Using the best available permit and guidelines information, we have been able to develop, for each of 45 parameters, a list of four-digit Standard Industrial Classification (SIC) codes ordered by the number of "major" discharges in each. There are a number of problems with this list: 1) "major" is defined by each EPA Region as it sees fit and there is not as uniform a measure of importance to NPDES as is number of permits or total discharge; 2) four-digit SIC codes may not differentiate waste types very well; 3) guidelines are not available for many SIC codes, particularly sewage treatment plants which account for almost half of all "major" dischargers. However, keeping its faults in mind, this list is the best tool currently available to assist EPA in specifying the sample types that must be studied before national approval of an alternative test procedure can be granted. It should be noted that option 3 (see Table 11) is feasible because the top five SIC codes include at least 45% of "major" dischargers for all parameters and includes 75% or better for all except BOD, pH and TSS. Option I would have only increased the BOD, pH and TSS coverage by 23%, 18% and 17% respectively, and this does not seem to justify the 233% cost increase".

TABLE 11. FEASIBLE DATA DESIGNS

Design	SIC codes	Samples SIC	Total analyses per sample	Total analyses	Cost ($12 ea)
1	10	10	8	800	$9,600
2	6	3	8	144	1,728 (-82%)
3	5	6	8	240	2.880 (-70%)

(From Reference 72).

This information sufficiently justifies the choice of various industrial discharge categories for NPDES method comparison. However, for individual parameters (e.g., BOD, suspended solids, pH and some nutrients), the largest discharge sources are municipalities. The lack of any municipal effluents in the data compilation will drastically affect the distribution of these selected parameters with respect to industrial discharges. Since a municipal effluent matrix can be significantly different from most industrial sources, the comparability of methods may be affected. Therefore, it seems appropriate to include a municipal effluent when certain parameters are being analyzed.

The EPA has recognized the potential problem and proposes to remedy it by a thorough evaluation of permit restrictions by parameter and type of discharge. Until the final information is available, EPA officials may include recommendations with respect to the type of industrial discharge to be sampled. If the parameter of interest is commonly specified in municipal permits, it may be appropriate to add or substitute a municipal effluent for a specified industrial effluent category. This determination can be made by EPA officials at the time of application.

SAMPLE STORAGE SYSTEM

As briefly presented in the previous section, obtaining samples may be the most costly part of a variance application. One method of overcoming this problem is to develop a central storage bank for sample storage and distribution. In general this approach was well received by QA coordinators, however, none desired the responsibility for maintaining this facility.

Limited storage space was the most commonly cited problem. In addition there is some question regarding the validity of this technique when unstable parameters are involved or when samples contain other unstable compounds. Therefore, all applicants will be required to collect and analyze their own samples. However, the specific industry samples should not be identified. The only identification needed will be by standard industrial classification (SIC) code.

EXPERIMENTAL DESIGN

There are certain factors which must be considered to allow a program for comparability testing to be successful. These include evaluation of a wide range of sample concentrations and minimization of variance associated with analysts and sample age. A well designed analytical method testing program would utilize a series of samples covering the range of concentrations of interest with the sample concentrations evenly dispersed between the low and high levels. The sequence of analyzing the samples and the sequence of performing the two methods being compared should be planned in advance to minimize the effect of the order of analyses on the results. If more than one analyst is involved, care should be taken that each analyst works on each method and on all types of samples. In order to obtain a good measure of variance for each method, only the prescribed number of replicate analyses should be performed and all the results should be reported.

It is recognized, however, that few applicants will be able to perform such a statistically well-designed experiment. For example, the applicant may try to obtain the required number of samples over a wide range of concentrations, but he has no assurance that the outfall being sampled will contain the desired concentrations during the sampling period available. It is also unlikely that the sample concentrations will be well distributed between the low and high levels. It is improbable that the applicant will utilize an experimental design that reduces the effect of extraneous variables (age of sample, analyst, etc) on the results. The person reviewing the results of the tests should also be aware that some applicants may be tempted to re-run an analysis when unexpected results are obtained and report only the "good" data, thus affecting the measure of variance.

Although the applicant may have little control over the concentrations of the samples he obtains, he can increase the range of concentrations tested by diluting and/or spiking portions of selected samples. He can also improve the distribution of the concentrations tested by blending portions of some samples. In this manner, a reasonable well-spaced, wide range of concentrations can be tested. It is not possible to set up a general fractional factorial experimental design suitable for each method comparison. The applicant can be advised, however, how to randomize the order of analysis and how to minimize the affect of operator influence on the method results. Failure to randomize these variables will tend to increase the chances of finding a significant difference between the two methods.

ESTIMATED COSTS OF A VARIANCE APPLICATION

The cost associated with preparing an application for use of an alternate procedure is a major consideration for prospective applicants. The use of a preliminary application form to indicate the scope of the requested variance and the submittal of a detailed description of the proposed method may make the performance of comparative tests unnecessary for every applicant. Previous comparative test data supplied by other applicants for a particular proposed method may be sufficient to accept an application for variance without further testing. If previous tests show that the proposed method is not comparable, however, the applicant should be allowed to submit comparative data anyway.

The costs of performing the method comparison tests can vary widely depending on the parameter measured, the type of variance requested, and the number and location of the sources to be tested. A laboratory applying for a permit-holder variance, for example, may be able to utilize samples obtained from the existing monitoring activity and have negligible sampling costs. A laboratory applying for a nationwide variance, however, may have to sample sources a considerable distance from the laboratory. This may require the addition of travel costs to the costs of sampling. To illustrate the costs that might be incurred, several examples have been prepared.

In these examples, the following assumptions have been made:

Time to obtain a 24 hour composite sample; includes set-up of sampler, compositing, clean-up, sample preservation, sample identification.	2 hours/sample
Time to obtain grab samples, includes sample preservation, identification, preparation	0.5 hour/sample
Travel time to sampling sites, local	2 hours/day
Travel time to sampling sites, distant	4 hours/day
Time required to prepare 6 additional samples by spiking, blending and diluting.	2 hours
Time required to complete data reports	1.0 hours/source
Sampling Labor	$16/hour
Analytical laboratory labor	$20/hour

The resultant costs as calculated are presented in Table 12.

Example 1 illustrates the expected cost when an applicant must sample only one source and can utilize samples from an existing monitoring activity. Example 2 shows the estimated cost for sampling three sources on the plant premises using grab samples. In Example 3, a laboratory must sample five industrial effluent types located within the same city as the laboratory. In Example 4, the five types of waste to be sampled are located at a greater distance from the laboratory. In this example, some of the analyses are performed by other laboratories which have higher rates to cover their own fee. In Example 5, the applicant subcontracts sampling and analysis to local laboratories to avoid the costs of travel and shipping. He must pay more for these services, however, than his own laboratory would charge.

The above examples are presented as illustrations only. It is obvious that the assumed times and pay scales are not universally applicable. Some additional costs, such as mileage charges, chemical costs, etc. have not been included.

DATA HANDLING

Evaluation of the various methods to handle the data using hand calculation, programmable calculators and computers was performed. The detailed Feasibility Study is included as Appendix B. The results indicate that a computerized data analysis and storage system would be needed to perform the necessary data manipulations. The proposed system is graphically outlined in Figure 5 and consists of three basic components:

TABLE 12. ESTIMATED SAMPLING COSTS FOR EQUIVALENCY TESTING

Example no.	1	2	3	4	5
Type of variance	Limited	Limited	Limited	Nationwide	Nationwide
Scope of variance	Permit-holder	Permit-holder	Laboratory	-	-
Sources tested	1	3	5	5	5
No. of samples obtained	6	18	30	30	30
No. of additional samples prepared	6	18	30	30	30
No. of analyses	14	144	240	240	240
Estimated times:					
Sampling	negligible	9 hrs	60 hrs	60 hrs	60 hrs[2]
Travel	-	-	60 hrs	120 hrs	-
Sample Prep.	2 hrs	6 hrs	10 hrs	10 hrs[1]	10 hrs[3]
Analysis	12 hrs	36 hrs	60 hrs	60 hrs[1]	60 hrs[3]
Reporting	1 hrs	3 hrs	5 hrs	5 hrs	5 hrs
Total	15 hrs	54 hrs	165 hrs	255 hrs	135 hrs
Estimated Costs	$300	$1,044	$3,420	$4,548	$2,932

1. 60% of analyses performed by another laboratory at 20% higher rate.
2. Samples obtained by local contractor at 20% higher rate.
3. All analyses performed by local laboratory at 20% higher rate.

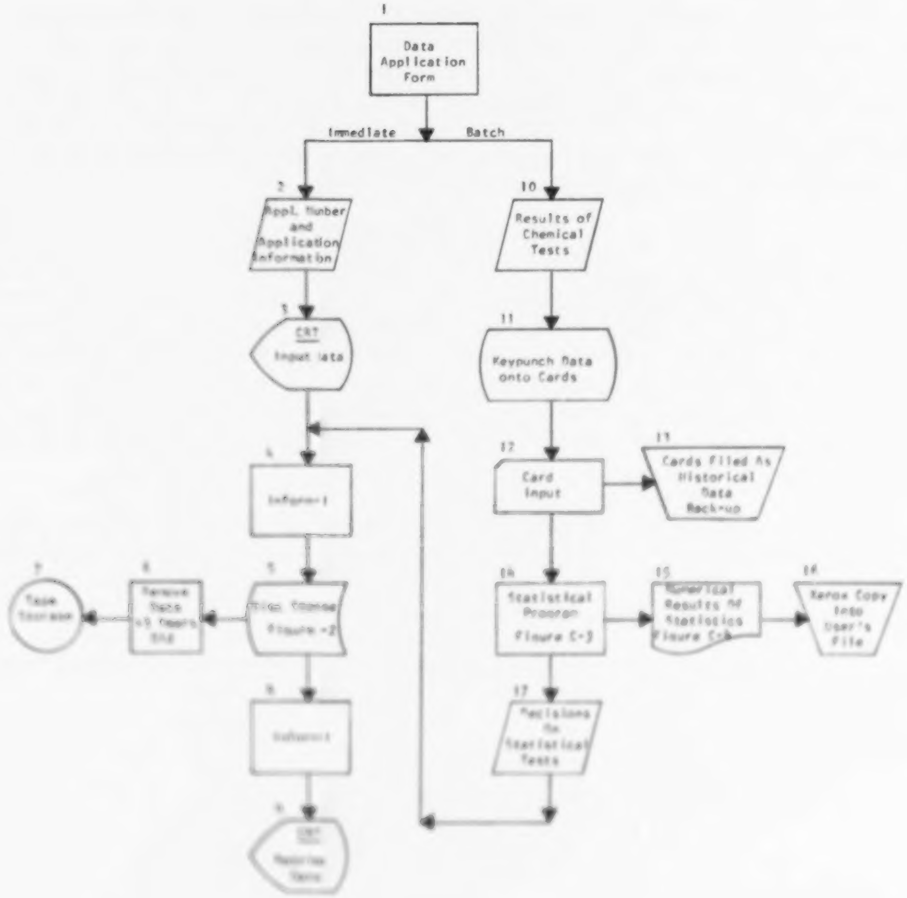

1
Data Application Form
Immediate
Batch
2
Appl. Number and Application Information
3
CRT
Input Data
4
5
Disc Storage Figure C-2
6
Remove Data >3 Years Old
7
Tape Storage
8
9
CRT
Retrieve Data
10
Results of Chemical Tests
11
Keypunch Data onto Cards
12
Card Input
13
Cards Filed As Historical Data Back-up
14
Statistical Program Figure C-3
15
Numerical Results of Statistics Figure C-4
16
Xerox Copy Into User's File
17
Decisions on Statistical Tests

Figure 8. Overview of the data analysis and storage system.

1. INFORM-II data handling system
2. Statistical program
3. Data storage

Data to be processed through the system is obtained from the equivalency data application form in the application packet. Background information concerning the variance request, such as applicant name and number, are processed immediately (10 seconds) through the Inform-II system. The results of the chemical tests are batch processed (one minute to daily turn around time) through the statistical program, which uses the previously discussed statistical analyses to test for data equivalency. Scatter plots, which visually present the equivalency data, are also produced as part of the output from the statistical program.

Data storage occurs at several locations within the system. Current application information is stored on disc and includes background applicant data and the statistical summary. This data can be readily retrieved using a cathode ray tube (CRT) terminal. Application data no longer considered current is removed from disc storage and stored on tape as historical data. Cards used to input data into the statistical program are also stored as historical data back-up. Numerical results of the statistical analyses performed by the statistical program are printed out, copied and stored in the applicants file.

An in-depth discussion of the data analysis and storage system, including the statistical analysis program, are in the Program Specification Report (Appendix C). This report presents sufficient detailed information to allow programming of the proposed computerized data handling system. The report presents the type of input required, the various calculations needed, and the type of output which can be anticipated from this approach. Detailed flow charts are also presented.

SECTION 8

DEVELOPMENT OF RECOMMENDED DATA EVALUATION PROCEDURE

SUMMARY OF EVALUATION

As a result of the total evaluation and the additional input provided by the implementation cosiderations, the main advantages and disadvantages of the statistical techniques were developed. From this information, the recommended protocol was established. All factors included in the total evaluation were considered in the final choice. This involved information drawn from the state-of-the-art review, the power of the statistics themselves, and the results of the data and cost evaluations.

Statistical Procedures

The various statistical procedures were evaluated in Section 6. A brief summary of the results is presented in the following paragraphs.

Outlier Rejection--

It is important to recognize that outlier rejection should occur only after careful consideration by the reviewer. Rejection of data points can significantly alter test results and give an improper indication of precision. Both techniques evaluated required an efficient estimate of the population standard deviation to be effective. This was the primary reason for four replicates in the EPA Interim Protocol. The efficiency of an estimate of the population standard deviation based on four observations is questionable at best. A more efficient estimate, i.e., one based on more replications would increase the cost of an application or reduce the number of sources evaluated. Given these alternatives, it is recommended that duplicates be analyzed instead of four replicates and that no formalized outlier rejection technique be applied to the observations.

Normality Testing--

The normalized histogram and the Moments test were evaluated for normality testing. The test results did not always agree. The Moments test was more sensitive to normality variations and gave more quantitative results since a number was calculated for comparison to a critical value. In this application, the number of samples was sufficient to perform a suitable evaluation of normality by method, and the Moments test is recommended.

Methods Comparison--

Data from different analytical methods were compared using the paired t-test, ANOVA and regression analyses. A discussion of these methods and results of comparative analyses appear in Section 6. All three tests offer different advantages and disadvantages which were confirmed or highlighted when data was applied. The paired t-test offers a simple, well known approach to establishing comparability but gives only a single number result. It is also less sensitive than ANOVA because data pairing is required. The ANOVA is the most sensitive of the three techniques. It also gives a single number result and the analyst can carefully evaluate the raw data for further information.

The regression analysis does not give a simple one result answer and, therefore may be more difficult to interpret. It is also weak when analyzing data which does not extend throughout the entire concentration range. But regression does provide much information about the relationship between the two methods. A knowledgeable analyst can evaluate the slope and intercept to establish if a systematic error has affected the data and check the "goodness of fit" of the line from the r^2 value. Both the autocorrelation coefficient and the Durbin-Watson result can alert the analyst to potential problems of error additivity or cross-over effects. Strata effects can also be interpreted using regression. Another advantage of regression analysis is that the technique has been commonly applied in method comparison studies (see Section 4). This indicates that chemists are familiar with the technique, understand the statistics involved and have accepted its use in this field.

Economic Considerations

It became apparent that the cost of comparative testing was dependent upon the number of analyses performed and the sampling costs. The number of analyses are affected by the number of industrial effluent categories evaluated, the number of samples and the number of replicates.

Industrial Effluent Categories--

The EPA Interim Protocol has recommended that five industrial effluent categories be evaluated, since for most parameters, five categories represent a majority of the parameter discharges. It has been recommended by the authors that a municipal effluent be added or replace an industrial effluent where appropriate. The specification of industrial and municipal effluent types to be analyzed for comparative testing should be established by EPA personnel, when a request for a variance is received.

Number of Samples and Replicates--

The EPA Interim Protocol requires six samples per industrial effluent category, which are analyzed in four replicates, for each method. This results in a total of 48 analyses per industrial effluent category. Both the paired t-test and regression use the mean of samples (average of replicates) to be utilized in the sensitivity of the statistical method. The ANOVA degrees of freedom are affected since the replicate values can be utilized.

The utilization of six samples has been shown to be cost-effective in terms of both analyses time and power of the statistics being utilized. Since in most cases, these six samples will not cover the range of concentration; mixing, blending, dilution and spiking within sample type by industrial effluent category could be used to cover a broader concentration range and provide a means of evaluating method accuracy.

RECOMMENDATIONS FOR PROTOCOL

Nationwide Variance

The final recommendation for nationwide variance includes the following steps and specifications:

Industrial Effluent Categories--
Five industrial effluent categories and a municipal category (where appropriate) should be specified by EPA personnel when an application is received. These should represent a large percentage of the discharges of the parameter being considered.

Number of Samples and Replicates--
Six samples of various concentrations should be collected from each industrial effluent category and analyzed by the approved and proposed technique in duplicate (i.e., analyses per method). Then the samples should be remixed within the individual discharge category so that they will provide data across the concentration range. Dilution, spiking or simple mixing of selected samples is recommended. Six additional samples should be prepared using those techniques and duplicate analyses should be performed on each sample by both analytical methods.

As a result of this approach, 48 analyses will have been performed for each industrial effluent category sampled, the same number as that required for the EPA Interim Protocol. However, a greater concentration range will be covered. This approach will also allow accuracy testing of the two methods.

Statistical Procedures Recommended--
It is recommended that the statistical evaluation of the test data include the following steps:

1. Calculation of basic statistics - The range and mean for each sample and source (industrial or municipal effluent category) are calculated. The sample ranges are compared using a paired t-test evaluation to determine which method is more precise. The calculation will involve subtraction of the proposed range from the approved and then calculation of a t value. Comparison against the 99% two tailed critical value will allow an evaluation of method precision.

2. Calculate the $\sqrt{b_1}$ and b_2 statistics for normality evaluation using the Moments test.[2] Critical values at the 99% confidence interval will be utilized for comparison.

3. Transform the data if necessary using one of three specified transforms. Transformation should begin with a log transform followed by a reciprocal or a square root transformation.

4. Perform a paired t-test comparison on the average of the duplicates and compare to the 95% two tailed critical value. Analyses should be performed by industrial effluent category, the total data, and on selected combinations.

5. Develop a scatter plot with the fitted regression line.

6. Calculate percent recoveries for accuracy evaluation.

This overall approach appears to offer the most information with respect to comparability. The basic statistics include the mean for the duplicates, as well as the range. The mean and range for the sources should also be calculated. The precision of the two methods can then be compared by calculating a paired t-test on the ranges. By consistently subtracting the proposed range from the approved, a correlation can be made with respect to which method is more precise. In addition, large ranges can alert the reviewer to potential precision problems.

Limiting the number of replicate analyses to two effectively eliminates the ability to statistically identify outliers. For this reason, outlier tests are not included in the recommended protocol. This approach appears to be justified since; 1) improper outlier rejection tends to bias results, and 2) it is expected that obvious outlier values may be replaced by data from analytical re-runs before the data are submitted for review. It should also be noted that no outliers were found in the data evaluated in Section 6.

Normality evaluation is recommended and the Moments test should be used. This test allows comparison of calculated statistics to critical values which makes it much more quantitative than the normalized histogram. If the data is shown to be non-normal, log, square root or reciprocol transformations can be applied. If these transformations do not result in a normal distribution, it is assumed that the data are non-parametric.

The reviewer can choose to reject the data, request more information or evaluate them with the understanding that a basic assumption has been violated. If one data set is normal and one is not, it is assumed that the data represent different populations and are, therefore, not comparable.

Both the paired t-test and regression have been included for method comparison. The paired t-test is a simple, straightforward evaluation which allows a single answer regarding method comparability. It could be used alone, but important information with respect to bias, strata, error additivity, etc., would not be obtained. For this reason, regression was included in the protocol. It is not as simple to interpret as the paired t-test, but it offers additional information which is useful in data evaluation. The use of two method comparison statistics gives the reviewer a more complete evaluation of the methods and should allow the best interpretation of the results.

The comparison of precision can be performed using a paired t-test on the ranges. An F test to evaluate precision will not be necessary, therefore, it has not been included.

Single Permit Holder Variance

The single permit holder variance evaluation would be identical to that above except that the applicant need only analyze for the discharges which are specified at a given location. In no instance will more than five locations be required. Multiple samples, plus sample mixing will still be necessary for each sampling location (source).

APPLICATION TO THE NIPDWR PROGRAM

In the course of the development of the procedure for testing the comparability of two analytical methods for the NPDES program, it became apparent that the procedure was also applicable to testing comparability for the NIPDWR program. Since EMSL-Cincinnati is involved in the technical review of applications for alternate analytical procedures for both programs, use of the same procedure for both programs is highly desirable.

It is expected that the statistical evaluation procedure developed for the NPDES program can be applied to the evaluation of comparative data from the NIPDWR program without modification.

SECTION 9

DEVELOPMENT OF USER'S DOCUMENTS

INTRODUCTION

Implementation of the recommended data evaluation procedure required the development of various user's documents which will allow consistent, efficient and effective evaluations of method comparability. It was determined that three different sets of information were required:

1. Orderly, clear and concise instructions for data collection and reporting for the applicant.
2. Suggestions for the consistent administration of the program, and guidelines.
3. Guidelines for those involved in data review.

These various documents were prepared and are attached to this report as Appendices D, E and F. The development of these packets involved many assumptions which are not part of their contents. Instead, the underlying approach and objectives have been presented in this section so the interested reader has a better understanding of the basis for their form and contents.

The data evaluation procedure and the development of the user's documents were originally intended for use in evaluating alternate methods for the NPDES system. It is expected that the data evaluation procedure can also be used to determine method equivalency for the NIPDWR system without modification. The implementation of the recommended sampling and test procedure and the associated user's documents, however, must accomodate the special requirements of the NIPDWR program and its participating agencies. The user's documents developed for the NPDES program are based on input from individuals from state and federal agencies involved in the NPDES program. The user's documents developed for NPDES program have been modified to reflect the author's understanding of the NIPDWR requirements based on the amendments to the regulations published in the Federal Register, July 9, 1976 (76) and the memorandum from Victor J. Kimm entitled, "Revised Procedure for Approval of Alternative Analytical Methods-- National Interim Primary Drinking Water Regulations", dated March 10, 1977 (77).

DEVELOPMENT OF APPLICANT USER'S PACKET

The applicant user's packet consists of two separate document types to be sent to the applicant by the appropriate agency. The first packet will be used to screen the general application before any comparative testing is performed. Then, only after a determination is made that comparative data is necessary, the second packet will be sent to the applicant with instructions pertinent to comparability testing requirements. Besides two types of packets, preliminary and detailed, the second form will differ depending on the type of variance requested.

Input from the various state and regional people greatly influenced the type of information included in the packet as well as its length. A sincere attempt was made to keep the application short and concise without omitting information critical to the applicant. An abbreviated application was developed which includes instructions on a line by line basis and an example of a completed form. This format was considered appropriate for direct computer input. In addition, the time frame for response was presented, as well as an official contact so the applicant could follow-up on the specific application at anytime. These inputs were in direct response to comments regarding the length of response time or the lack of knowledge of the status of an application for extended periods.

Preliminary Application Form

Upon written request for a variance application, the state or regional administrator is to send a preliminary application to the applicant (see Appendix D). This form is general and applicable to all types of variance requests.

The purpose of this application is to screen the requests prior to the performance of comparability testing. This screening may reduce (or even eliminate) the amount of comparability testing needed. There are several circumstances which could merit this type of screening:

1. A similar method has been evaluated and granted a variance already. Therefore, further testing may be unnecessary. Or, similarly, sufficient data may be on file which will allow evaluation of the method based on comparative tests on only one or two waste sources.

2. The proposed procedural difference is only a minor change in the approved technique, so a variance is unnecessary.

3. The procedure has already been evaluated and found non-comparable. Informing the applicant of prior findings of the non-comparability of the proposed method may save the applicant the cost of the comparability testing. This notification of non-comparability, however, should not prevent the applicant from performing the tests and submitting his own comparability data for a re-evaluation of the proposed method.

In the packet, the applicant is asked to submit basic data on the application, as well as a detailed, referenced description of the proposed comparable technique. The applicant is also asked to identify the type of variance desired with regard to the NPDES or NIPDWR programs. Both limited use or nationwide variances are possible. It is expected that applications for nationwide variances will be forwarded to EMSL-Cincinnati for review. Applications for limited variances will probably be considered by the regional offices with technical help from EMSL-Cincinnati.

If the information in the preliminary application is insufficient to grant the requested variance, the appropriate detailed application packet should be sent to the applicant. This packet identifies the waste or water sources to be sampled, instructs the applicant on how to perform the comparability tests, and includes data reporting forms.

Detailed Equivalency Application Packet

If it is determined that additional information is needed, the applicant can be sent the appropriate Detailed Equivalency Application Packet.

Four different types have been prepared: 1) NPDES parameter, nationwide, 2) NPDES parameter, limited use, 3) NIPDWR, nationwide, and 4) NIPDWR, limited use. These have been developed separately so that the applicant will not be required to sort through the application to find the appropriate sections. The administrator can refer to the preliminary application form to determine which type is appropriate.

Depending on the type of variance requested, the person sending out the detailed equivalency application packet should include two or three types of information to the applicant. For all packets, the sources of waste or water to be tested by the applicant should be listed on the first page of the instructions. In addition, the application identification numbers assigned to the preliminary form should be entered on each data reporting application. For the NIPDWR limited use packet, the name and address of the agency which should receive the completed data forms should also be listed on the first page. For this type of variance, the appropriate agency depends on the scope of the limited variance request and the State of residence of the applicant. The detailed packets for the other types of variance requests include the name of the applicable agency.

As indicated previously, an EPA person to contact and an estimated time for application review are included. Sampling and analytical directions are presented to aid the applicant. The most complicated aspect of these directions, associated with the dilution, spiking and blending of samples, is explained in detail and a special table is included as an example.

The total number of samples required is summarized so the applicant can determine if the directions are consistent with his/her understanding. Finally, some recommendations are made regarding ways to minimize systematic variations in the data which could adversely affect a comparability evaluation. This type of information is intended to help the laboratory submit the data which most closely approximates the assumptions required for the statistics.

Data reporting directions are presented in detail. An example of a completed form is included for reference. Finally, the applicant is instructed as to where the data should be sent and how long it will be before the evaluation of the application is complete.

A brief description of the statistical evaluation procedure is also included. It is felt that the applicant needs to be aware of the type of evaluation being performed. In the interest of space, critical values and other details have not been included. Applicants interested in more details can contact the Director of EMSL in Cincinnati.

DEVELOPMENT OF GUIDELINES FOR PROGRAM ADMINISTRATION

In order to administer the program more consistently and to aid those personnel involved with receipt of variance requests, suggested guidelines have been prepared. These guidelines are addressed to either the Regional office or EMSL - Cincinnati.

These guidelines are suggestions for the implementation of the equivalency program based on the author's understanding of the inter-relationship of the federal and state agencies involved in the process of reviewing variance applications.

Guidelines for the Regional Office

The information presented includes identification of responsibility with respect to the person granting the variance. The Regional Administrator has authority to grant limited use NPDES or limited use NIPDWR variances after consideration of the technical evaluation performed by EMSL. The guidelines also include the agency where applications should be forwarded after receipt by the Regional Administrator.

Suggested procedures are included for the review of limited use variance procedures by the Regional Administrator. Review involves an initial evaluation of the preliminary application to determine if the applicant should be requested to submit comparability data. It is the opinion of the authors that data should only be submitted after this preliminary review to minimize unnecessary comparability testing by the applicant.

If data is required, suggested methods to identify water or waste sources are presented. A regional administrator is strongly encouraged to contact EMSL-Cincinnati if questions regarding source specifications arise.

Application and file numbers should also be specified by the administrator. Since the applicant may be calling the administrator to determine status of a specific application, the assigned number should be suitable for the applicant to use it as a reference. Prearranged procedures for assigning application and file numbers are desirable.

Finally, it will be necessary for the Regional Administrator to complete the status report at the bottom of Form 1 - Preliminary Application. In this way, inquiries regarding status can be responded to accurately.

There are no guidelines given for actual review of the results. A technical evaluation and recommendation will be made by the EMSL staff. Other decision criteria are up to the administrator, however, all parties concerned should be notified of the final decision.

Guidelines for EMSL - Cincinnati

The technical evaluation of all request for variance along with an associated recommendation to the appropriate party is made by EMSL-Cincinnati.

Once a preliminary application for either NPDES or NIPDWR nationwide variance is received, it should be reviewed to determine if additional data must be submitted by the applicant. These considerations are presented in the attached guidelines. As indicated previously, the authors feel that collection of unnecessary data prior to the review could involve a needless expense for the applicant.

After determination that more data is necessary, the appropriate application and file numbers must be assigned to the preliminary application form and the status report completed. The appropriate detailed equivalency application packet should be prepared by specifying the waste or water sources to be sampled and adding the application numbers to the data reporting form. The packet should then be sent to the applicant for the method comparison tests.

Once the completed data is received, the EMSL technical staff can perform the proposed computerized evaluation. The resultant recommendation is sent to the appropriate authority for the approval or rejection of the variance application. Once the final decision has been reached, it is suggested that all agencies involved be informed.

DEVELOPMENT OF GUIDELINES FOR THE REVIEW OF COMPARATIVE TEST DATA

The reviewers guidelines are intended to help the technical reviewers evaluate the results of the statistical analysis of the comparative test data. These guidelines are written for the use by a reviewer who is not a statistician.

The first step in the review is to check whether the applicant followed the instructions for the type of variance requested. The sources sampled, the number of samples obtained from each source, the approved method used for comparison and the laboratories performing the analyses should all be checked.

The reviewer should also check the data for obvious inconsistencies in data format. The reviewer can quickly resolve questions concerning the reported data by a telephone call to the name and number listed on line 6 of the data report forms.

After the data report forms have been checked, the data should be analyzed using the computer program developed for this comparison. The second portion of the Reviewer Guidelines describes the expected computer printout, and presents information intended to aid in the interpretation of the printout. Options available to the reviewer and possible causes of inconsistencies in the results are indicated.

SECTION 10

REFERENCES

1. Sladek, K. J., et. al., Statistical Interpretation of Membrane Filter Bacteria Counts, Millipore Corporation, pp. 1-14, January 10, 1975.

2. Schaeffer, D. J., et. al., Statistical Analysis of the Recovery of Coliform Organisms on Gelman and Millipore Membrane Filters, Applied Microbiol., pp. 605-607, October 1974.

3. Stuart, D. G., et. al., Membrane Filter Techniques for the Quantification of Stressed Fecal Coliforms in the Aquatic Environment, Applied and Environ. Microbiol., 34 (1):42-46, July 1977.

4. Green, B. L., et. al., Two Temperature Membrane Filter Method for Enumerating Fecal Coliform Bacteria from Chlorinated Effluents, Applied and Environ, Microbiol., 33 (6):1259-1264, June 1977.

5. Recent Results from In-House Ozone Contactors Disinfection Study, Technology Transfer, U.S. Environmental Protection Agency, October 1977.

6. Lin, Shundar, Evaluation of Fecal Streptococci Tests For Chlorinated Secondary Sewage Effluents, Journal of the Environmental Engineering Divison, pp. 253-267, April 1974.

7. Green, B. L., et. al., Comparison of New Millipore HC with Conventional Membrane Filters for the Enumeration of Fecal Coliform Bacteria, Applied Microbiol., 31(4):697-699, 1975.

8. Lin, S. D., Comparison of Membranes for Fecal Coliform Recovery in Chlorinated Effluents, Journal WPCF, 49(11):2255-2264, November, 1977.

9. Verstraete, V., et. al., Three Step Measurement by the Sapromat to Evaluate the BOD$_5$, the Mineral Imbalance and the Toxicity of Water Samples, Water Research, 8:1077-1081, 1974.

10. Winter, J. A., Method Research Study 3, Demand Analyses, - An Evaluation of Analytical Methods for Water and Wastewater, Environmental Protection Agency, pp. 9-22, 1971.

11. Tebbutt, T. H. Y., et. al., Respirometric Determination of BOD, Water Research, 10:613-617, 1976.

12. Tuffey, T. J., et. al., A Critical Analysis of Warburg Respirometry for BOD Determinations of Polluted Waters, Proceedings of the 29th Industrial Waste Conference, pp. 1-8, May 7, 8, and 9, 1974.

13. Salotto, Vincent B., et. al., Elemental Analysis of Wastewater Sludges From 33 Wastewater Treatment Plants in the United States, Research Symposium on Pretreatment and Ultimate Disposal of Wastewater Solids, Rutgers University, New Brunswick, New Jersey, pp. 8-14, May 21-22, 1974.

14. Yeager, David W., et. al., Determination of Lead in Biological and Related Material by Atomic Absorption Spectrophotometry, Environmental Science & Technology, 5:1020-1022, October 1971.

15. Jackson, Danny R., et. al., CD Ion Activity in Soils: Evaluation of CD Ion-Selective Electrode, Environmental Science & Technology, 11(6):585-587, June 1977.

16. Raman, A., A Comparison of Two Methods for the Determination of Serum Ionic Calcium, Clin. Biochem., 5:208-213, 1972.

17. Ostling, Goran, Short Communication, - A Simplified Automated Curcumin Method For the Determination of Boron in Sea Water, Analytical Chemical Acta., 78:507-512, 1975.

18. Lenz, Bernard L., et. al., Ion-Selective Electrode Method Compared to Standard Methods for Sodium Determination in Mill Liquors, TAPPI, pp. 2051-2055, December, 1971.

19. Winter, J. A., et. al., Interlaboratory Study of the Cold Vapor Technique for Total Mercury in Water, Water Quality Parameters, ASTM STP 573, pp. 566-580, 1975.

20. Robbins, D. A., et. al., Results of Trace Arsenic Analyses Performed by Various Laboratories, American Industrial Hygiene Association Journal, pp. 57-60, January, 1977.

21. Yu, Kar Y., et. al., Evaluation of a Nitrate-Specific Ion Electrode, Journal WPCF, 49(8):1896-1901, August, 1977.

22. Winter, J. A., et. al., An Evaluation of Analytical Methods for Water Wastewater, Environmental Protection Agency, - Method Study 2, Nutrient Analysis Manual Methods, pp. 10-46, 1970.

23. Johnson, David L., Simultaneous Determination of Arsenate and Phosphate in Natural Waters, Environmental Science & Technology, 5(5):411-414, May, 1971.

24. Jirka, Andrea, et. al., Ultramicro Semiautomated Method For Simultaneous Determination of Total Phosphorus and Total Kjeldahl Nitrogen in Wastewaters, Environmental Science & Technology, 10(10):1038-1044, October, 1976.

25. Senn, David R., et. al., Determination of Nitrate Ion at the Part Per Billion Level in Environmental Samples With a Continuous Flow Immobilized Enzyme Reaction, Analytical Chemistry, 48(7):954-958, June, 1976.

26. Osburn, Quentin W., et. al., Automated Method for Ortho-, Ortho-Plus Hydrolyzable, and Total Phosphate in Surface and Wastewaters, Environmental Science & Technology, 8(4):363-366, April, 1974.

27. Thomas, Robert F., Selective Electrode Measurement of Ammonia in Water and Wastes, Environmental Science & Technology, 7(6):523-526, June, 1973.

28. Bietz, J. A., Micro-Kjeldahl Analysis by an Improved Automated Ammonia Determination Following Manual Digestion, Analytical Chemistry, 46(11):1617-1618, September, 1974.

29. Bremner, J. M., Use of an Ammonia Electrode for Determination of Ammonium in Kjeldahl Analysis of Soils, Comm. in Soils Science & Plant Analysis, 3(2):159-165, 1972.

30. Katsuhika, Matsunaga, A Rapid and Sensitive Method for Determination of Submicrogram Amount of Ammonia in Fresh and Sea Waters, Analytica Chimica Acta., 73:204-208, 1974.

31. Carter, M. J., and Huston, M. T., Preservation of Phenolic Compounds in Wastewaters. Environmental Science & Technology, 12(3):309-313, March, 1978.

32. Myers, L. H., et. al., Analytical Variability of Five Wastewater Parameters - Petroleum Refining Industry, EPA-600/2-76-234, pp. 1-67, September, 1976.

33. Fountaine, J. E., New Ultraviolet Ratio Spectrophotonetric System For the Determination of Trace Amounts of Phenolic Compounds, Analytical Chemistry, 46(1):62-66, January, 1974.

34. Beasley, M. D., et. al., Sampling Errors in the Quantitation of Petroleum in Boston Harbor Water, Analytical Chemistry, 46(12):1858-1860, October, 1974.

35. Hofton M., Continuous Determination of Free Cyanide in Effluents Using Silver Ion Selective Electrode, Environmental Science & Technology, 10(3):277-280, March, 1976.

36. Fine, D. H., et. al., Determination of Dimethylnitrosamine in Air and Water by Thermal Energy Analysis: Validation of Analytical Procedures, Environmental Science & Technology, 11(6):577-580, June, 1977.

37. Crisp, P. T., et. al., The Determination of Anionic Detergents With the bis (Ethylenediamine) Copper (11) Ion, Analytica Chimica Acta., 78:391-396, 1975.

38. Afghan, B. K., et. al., An Improved Method for the Determination of Trace Quantities of Phenols in Natural Waters, Analytica Chimica Acta., 71:355-366, 1974.

39. Gales, M. E., An Evaluation of the 3- Methyl -2 - Benzothiazolinone-Hydrazone Method for the Determination of Phenols in Water and Waste Waters, Analyst, 100:841-847, December, 1975.

40. Winter, J. A., et. al., An Evaluation of FWPCA Analytical Methods For Surface Waters, FWPCA Method Study 1 Mineral and Physical Analyses, pp. 1-33, June, 1969.

41. Selmer-Olsen, A. R., Determination of Chloride in Aqueous Soil Extracts and Water Samples by Means of a Chloride - Selective Electrode, Analyst, 98:412-415, June, 1973.

42. Chen, K., et. al., Solvent Extraction of Sulfur From Marine Sediment and its Determination by Gas Chromatography, Environmental Science & Technology, 7(10):948-951, October, 1973.

43. Hashimoto, Y., et. al., Preparation of Homogeneous Dust Sample For Comparison of Analytical Results of Atmospheric Samples, Environmental Science & Technology, 10(8):815-818, August, 1976.

44. MacLeod, K. E., et. al., Comparison of the Spadns - Zirconium Lake and Specific Ion Electrode Methods of Flouride Determination in Stack Emission Samples, Analytical Chemistry, 45(7):1272-1273, June, 1973

45. Devine, R. F., et. al., Interference of Sulfate Ion on Spadns Colorimeteric Determination of Flouride in Waste Waters, Environmental Science & Technology, 9(7):678-679, July, 1975.

46. Crosby, N. T., et. al., An Evaluation of Some Methods for the Determination of Flouride in Potable Waters and Other Aqueous Solutions, Analyst, 93:643-652, October, 1968.

47. Robertson, D. J., et. al., Improvements in Phenoldisulfonic Acid Method for Determination of NO_x , Environmental Science & Technology, 9(10):979-980, October, 1975.

48. Finger, R. E., et. al., An Automated Spectrophotometric Suspended Solids Analysis for Activated Sludge, Journal WPCF, 47(5):1043-1054, May, 1975.

49. A Preliminary Review of Analytical Methods For the Determination of Suspended Solids in Paper Industry Effluents for Compliance with EPA-NPDES Permit Terms, National Council of the Paper Industry for Air and Stream Improvement, Inc., Special Report #75-01, pp. 7-10, January, 1975.

50. Babcock, R. H., et. al., Selective Ion Electrode System for Flouride Analysis, Journal AWWA, pp. 953-961, August, 1968.

51. Kanipe, L. G., Handbook for Analytical Quality Control in Radiological Laboratories, E-EP/77-4, EPA-600/7-77-088, August, 1977.

52. Green, Arthur C., et. al., Development of a System for Conducting Inter-Laboratory Tests for Water Quality and Effluent Measurements, EPA-600/4-77-031, pg. 7, June, 1977.

53. Anscombe, F. J., et. al., The Examination and Analysis of Residuals, Technometrics, 5(2):141-159, May, 1963.

54. 1977 Annual Book of ASTM Standards, Determination of Precision and Bias of Methods of Committee D-19 on Water, Part 31, pp. 7-15.

55. Youden, W. J., and Steiner, E. H., Statistical Manual of the Association of Official Analytical Chemists, Association of Official Analytical Chemists, pp. 1087, 1975.

56. Grubbs, F. E., Procedures for Detecting Outlying Observations in Samples, Technometrics, 11(1):1-21, February, 1969.

57. Shapiro, S. S., et. al., A Comparative Study of Various Tests for Normality, American Statistical Association Journal, pp. 1343-1372, December, 1968.

58. Duncan, A. J., Quality Control and Industrial Statistics, Richard D. Irwin, Inc., pp. 412-417, 1952.

59. Bennett, C. A.; and Franklin, N. L., Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons, Inc., pg. 620, 1954.

60. Ostle, B., Statistics in Research, The Iowa State University Press, pg. 471, 1963.

61. Communication with Dr. G. Liberson, Washington, EPA.

62. Bartlett, M. S., The Use of Transforms, Biometrics, 3:39-51, 1947.

63. Davies, O. L., The Design and Analysis of Industrial Experiments, Hafner Publishing Co., pg. 45, 1954.

64. Hajek, J., Nonparametric Statistics, Holden Day, pp. 52-56, 1969.

65. Miller, I., and Freund, J. E., Probability and Statistics for Engineers, Prentice-Hall, Inc., pp. 170, 215, 1965.

66. Walsh, J. E., Applications of Some Significance Tests for the Median Which are Valid Under Very General Conditions, Journal American Statistical Association, 44(2,3):355-442, September, 1949.

67. Dixon, W. J., and Massey, Jr., F. J., Introduction to Statistical Analysis, McGraw-Hill Book Company, Inc., New York, New York, pg. 252, 1951.

68. Youden, W. J., Statistical Manual of the Association of Official Analytical Chemists, Statistical Techniques for Collaborative Tests, Association of Official Analytical Chemists, pp. 21-29, January, 1975.

69. Hilton, G., Intermediate Pilitometrics, Columbia University Press, New York, 1976.

70. Concepts and Applications of Regression Analysis, IBM Data Processing Application, Technical Publications, Dept., White Plains, N. Y., pg. 29, 1966.

71. Daniel, C. and Wood, F. S., Fitting Equations to Data, Wiley Inter-Science, John Wiley & Sons, 5:207, September, 1949.

72. Bartlett, M. S., Fitting a Straight Line When Both Variables are Subject to Error, Biometrics, pg. 207, September, 1949.

73. Cochran, W. G., Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied, Biometrics, 3:22-38, 1947.

74. Letter from Mr. Paul W. Britton, Statistician to the Record, June 16, 1977, Environmental Monitoring and Support Laboratory - Cincinnati, USEPA.

APPENDIX A
Literature and Data Used in Review of Present Practices and Evaluation Statistical Methods

TABLE A-1. SUMMARY OF ANALYTICAL COMPARISON METHODS REVIEWED IN THE LITERATURE

Reference	Evaluate distribution	Transform	Spiked recovery	Method comparison	Analyte matrix
Bacteriology					
B-1	observation	log	no	regression	water
B-2	observation	log	no	regression	water
B-3	no	no	no	ANOVA	water
B-4	observation	log	no	regression	water
B-5	observation	log	no	Duncan's multiple range	water
B-6	observation	log	no	t-test	water
B-7	observation	log	no	ANOVA	water
B-8	no	no	no	t-test regression	water
B-9	no	no	no	observation	water
B-10	no	no	no	observation	water
Demand					
D-1	histogram, skewness	no	yes	observation, t-test	TOC - 91 COD - 12 BOD - 10

(continued)

TABLE A-1. (continued)

Reference	Evaluate distribution	Transform	Spiked recovery	Method comparison	Analyte matrix
Demand (continued)					
D-2	no	no	yes	mean ratio	COD - 12
D-3	no	log	yes	regression	BOD - 10
D-4	no	no	yes	mean ratio	COD - 12
D-5	no	no	yes	regression	BOD - 10
D-6	no	no	yes	F-test	BOD - 10
D-7	no	no	yes	observation	(solid waste) carbon
Metals					
M-1	histograms	log	yes	observation	sludges
M-2	no	no	no	ANOVA	paper mill wastewater
M-3	observation	log	yes	regression	biological serum
M-4	observation	log	yes	regression	soils
M-5	no	no	yes	Youden's paired samples	water
M-6	no	no	no	regression, t-tests	bio-serum
M-7	no	no	yes	t-tests	air

(continued)

TABLE A-1. (continued)

Reference	Evaluate distribution	Transform	Spiked recovery	Method comparison	Analyte matrix
Metals (continued)					
M-8	no	no	no	t-tests	water
M-9	no	no	no	regression	sea water
M-10	no	no	yes	observation	sea water
M-11	no	no	yes	observation	water
M-12	no	no	yes	observation	water
M-13	no	no	yes	observation	water
M-14	no	no	yes	observation	water
M-15	no	no	yes	observation	waters and wastewaters
M-16	no	no	yes	observation	water
M-17	no	no	yes	observation	waters
M-18	no	no	yes	observation	waters
M-19	no	no	yes	observation	waters
M-20	no	no	yes	Chi squared	air

(continued)

TABLE A-1. (continued)

Reference	Evaluate distribution	Transform	Spiked recovery	Method comparison	Analyte matrix
Nutrients					
N-1	coef. of skewness	no	yes	Youden's paired samples	NH_3 - 3 NO_3 - 88 o-PO_4 -93 TKN - 23 t-PO_4 -98
N-2	(Reject outliers by Dixon's test)	no	no	regression	PO_4 - (seawater)
N-3	no	no	no	regression	NH_3 - 3
N-4	observation	log	yes	ANOVA t-test	NO_3 - 88
N-5	no	no	yes	regression t-test	t-PO_4 -98 TKN - 23
N-6	no	no	yes	regression	NO_3 - 88
N-7	no	no	yes	regression	TKN - soils
N-8	no	no	yes	regression	NH_3 - 3
N-9	no	no	no	observation	NO_3 - 88

(continued)

TABLE A-1. (continued)

Reference	Evaluate distribution	Transform	Spiked recovery	Method comparison	Analyte matrix
Nutrients (continued)					
N-10	no	no	yes	observation	PO_4-P(all forms)
N-11	no	no	yes	% difference	NH_3 - 3
N-12	no	no	no	observation	NO_3 - 88
N-13	no	no	no	observation	o-PO_4 - 98
N-14	no	no	yes	observation	NO_3 - soil
Organics					
O-1	no	no	yes	F-test t-test	Oils - 90 COD - 12 SS - 106 phenol - 96
O-2	no	no	yes	t-test	phenols - 96
O-3	no	no	yes	observation	phenols - 96
O-4	no	no	yes	observation	Oils - water
O-5	no	no	yes	observation	organics in air
O-6	no	no	yes	observation	CN - 17

(continued)

TABLE A-1. (continued)

Reference	Evaluate distribution	Transform	Spiked recovery	Method comparison	Analyte matrix
Organics (continued)					
0-7	no	no	yes	observation	detergents in water
0-8	no	no	yes	observation	phenols - 96
0-9	no	no	yes	observation	phenols - 96
0-10	no	no	no	observation	organics - water
0-11	no	no	no	observation	organics - water
Physical, Mineral, Residue					
P-1	no	no	yes	Youden's paired sample	pH - 22, TDS - 105, sp. cond. - 109, hardness - 21, Na - 76, K - 67, acidity/alk. - 1,2, Cl^- -13, SO_4 - 110
P-2	no	no	no	regression	S. S. - 106
P-3	no	no	yes	t-test	Cl^- - soil
P-4	no	no	no	regression with ANOVA	SS - activated sludge

(continued)

TABLE A-1. (continued)

Reference	Evaluate distribution	Transform	Spiked recovery	Method comparison	Analyte matrix
Physical, Mineral, Residue (continued)					
P-5	no	no	yes	t-test	S - sediments
P-6	no	no	yes	F-test	Al, Va, Cl_2 - dust
P-7	no	no	no	regression	F - 20
P-8	no	no	yes	regression	F - stack samples
P-9	no	no	no	observation	F - air
P-10	no	no	no	ratio of means	$SO_4^=$ - air
P-11	no	no	yes	observation	F - 20
P-12	no	no	no	observation	$SO_4^=$ - air
P-13	no	no	yes	observation	F - 20
P-14	no	no	no	observation	S^- - 111
P-15	no	no	no	observation	$SO_4^=$ - water
P-16	no	no	yes	observation	NO_x - air
P-17	no	no	no	observation	MBAS - waters
P-18	no	no	no	observation	F - 20
P-19	no	no	no	observation	Cl^- - bioserum

(continued)

TABLE A-1. (continued)

Reference	Evaluate distribution	Transform	Spiked recovery	Method compariton	Analyte matrix
Radiological					
R-1	no	no	no	Chauvenet's outlier method	seaweed
R-2	no	no	no	Chauvenet's	water, seaweed
R-3	observation	no	yes	observation	water

TABLE A-2 . RAW DATA AND BASIC STATISTICS FOR COD-A AND COD-B

I COD-A Data

Mining (uranium mill wastes)

			Sample			
Method	1	2	3	4	5	6
Approved	64	28	65	71	24	16
	63	27	65	71	26	12
	63	23	64	73	20	18
	60	21	63	69	24	20
$x^{a.}$	62	25	64	71	24	17
$s^{b.}$	1.71	3.30	0.96	1.63	2.52	3.41
Proposed	66	26	56	69	22	20
	62	25	56	72	31	18
	61	19	56	69	18	22
	66	25	62	71	28	15
X	64	24	58	70	25	19
S	2.63	3.20	3.00	1.50	5.86	2.99

Petroleum (refining)

			Sample			
Method	1	2	3	4	5	6
Approved	30	28	41	34	65	65
	28	30	36	33	102	69
	33	29	40	36	81	67
	33	28	38	36	83	63
X	31	29	39	35	83	66
S	2.45	0.96	2.22	1.50	15.15	2.58
Proposed	29	29	40	38	80	68
	25	32	34	40	99	63
	29	32	38	34	95	66
	32	27	41	32	89	66
X	29	30	38	36	91	66
S	2.87	2.45	3.10	3.65	8.26	2.06

a. Arthimetic mean.
b. Standard deviation.

(continued)

TABLE A-2. (continued).

Inorganic chemical (manufacturing)

	Sample					
Method	1	2	3	4	5	6
Approved	752	1024	3248	80	912	952
	752	1024	3264	80	912	952
	752	1020	3248	80	912	952
	748	1024	3264	80	928	960
X	751	1023	3256	80	916	954
S	2.00	2.00	9.23	0.00	8.00	4.00
Proposed	798	1026	3392	68	908	968
	794	1034	3392	68	912	968
	798	1022	3392	70	908	950
	794	1030	3392	70	912	946
X	796	1028	3392	69	910	958
S	2.31	5.16	0.00	1.15	2.31	11.66

Plastics industry

	Sample					
Method	1	2	3	4	5	6
Approved	64	84	1168	1372	38	52
	62	80	1160	1364	48	48
	64	84	1164	1352	52	48
	64	80	1160	1352	38	44
X	65	82	1163	1360	44	48
S	1.00	2.31	3.83	9.80	7.12	3.27
Proposed	74	97	1140	1368	48	59
	72	97	1160	1368	44	55
	76	97	1148	1368	49	55
	76	93	1148	1352	48	57
X	75	96	1149	1364	47	57
S	1.91	2.00	8.25	8.00	2.12	1.91

(continued)

TABLE A-2. (continued).

Textile wastes

	Sample					
Method	1	2	3	4	5	6
Approved	268	264	629	363	355	192
	268	268	625	361	357	192
	270	268	627	363	355	192
	270	268	629	361	355	192
X	269	267	628	362	356	192
S	1.54	2.00	1.91	1.15	1.00	0.00
Proposed	270	268	635	382	374	190
	274	270	636	376	372	192
	268	266	636	374	376	186
	268	268	635	372	374	190
X	270	268	635	376	374	190
S	2.83	1.63	0.58	4.32	1.63	2.52

II COD-B Data

Textile wastes

	Sample					
Method	1	2	3	4	5	6
Approved	34	76	35	47	46	31
	34	115	47	47	39	31
	49	73	39	50	39	23
	42	80	43	50	39	27
X	39	86	41	48	41	28
S	7.22	19.54	5.16	1.73	3.50	3.82
Proposed	42	95	46	52	38	38
	45	93	69	69	37	41
	40	88	78	53	40	68
	39	105	49	61	40	40
X	42	95	61	59	39	47
S	2.64	7.13	15.50	7.93	1.50	14.22

(continued)

TABLE A-2. (continued).

Domestic wastes

			Sample			
Method	1	2	3	4	5	6
Approved	66	19	49	191	108	89
	66	27	61	191	112	85
	69	31	57	195	108	93
	69	27	53	187	93	93
X	67	26	55	191	105	90
S	1.73	5.03	5.16	3.26	8.38	3.82
Proposed	73	37	62	181	83	72
	81	42	46	181	115	87
	70	41	51	174	112	69
	72	41	52	230	86	95
X	74	40	53	192	99	81
S	4.83	2.21	6.70	25.87	16.83	12.33

Hospital effluent

			Sample			
Method	1	2	3	4	5	6
Approved	85	78	78	70	86	66
	101	74	82	62	86	66
	97	74	90	55	86	66
	93	82	82	59	86	66
X	94	77	83	62	86	66
S	6.83	3.82	5.03	6.35	0	0
Proposed	100	88	90	34	93	99
	70	94	118	33	91	95
	77	129	123	34	91	120
	64	122	128	50	111	95
X	78	108	115	38	97	102
S	15.75	20.27	16.99	8.18	9.59	11.98

(continued)

TABLE A-2. (continued).

Dye industry

	Sample					
Method	1	2	3	4	5	6
Approved	143	228	105	101	86	125
	154	220	105	101	78	125
	131	224	128	121	82	121
	154	216	120	105	78	121
$\bar{X}$	145	222	114	107	81	123
S	10.96	5.16	11.44	9.52	3.82	2.30
Proposed	127	222	89	109	52	153
	112	218	90	112	52	123
	107	190	87	118	61	153
	125	195	92	110	52	116
$\bar{X}$	118	206	89	112	55	136
S	9.77	16.58	2.08	4.03	4.50	19.55

TABLE A-3. EXPECTED ERROR USING NPDES ANALYTICAL METHODS.

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Alkalinity	2	Electrometric pH 4.5	EPA	Syn. water	8 mg/l	42	**
					9	48	**
					113	18	*
			APHA		114	16	*
					120	16	*
Ammonia	3	Manual distillation, nesslerization	APHA	Dist. water	200 µg/l	33-102	
					800	36-51	
					1,500	13-40	
		Manual distillation, titration			200	160	***
					800	62	***
					1,500	47	**
		Automated phenolate			200	47	**
					800	35	**
					1,500	15	*
		Electrode	EPA	Surface water	0.13 mg/l	9	*
					0.19	8	*
					0.77	40	**
					1.00	14	*
Biochemical oxygen demand (BOD)	10	Winkler or Electrode	APHA	N.G.	194 mg/l	37	**
Chemical oxygen demand (COD)	12	Dichromate reflex,	EPA	Dist. water	270 mg/l	17	*
		Dichromate reflex diluted method			12.3	68	***
Chloride	13A	Silver nitrate	APHA	Dist. water	241 mg/l	10	*

Method Rating
- * Excellent 0-25%
- ** Acceptable 25-50%
- *** Unacceptable >50%

(continued)

TABLE A-3 (continued).

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
	13B	Mercuric nitrate	EPA	Syn. water	17 mg/l	20	*
					13	18	*
					91	7	*
					97	7	*
					382	7	*
					398	7	*
			APHA	Dist. water	241	10	*
	13C	Automated colorimetric ferricyanide	EPA	Surface water	100 mg/l	5	*
Chlorine	15A	Iodometric titration	APHA	Dist. water	640 µg/l	83	***
					840	73	***
					1,830	64	***
	15B	Amperometric	APHA	Dist. water	640 µg/l	58	***
					1,830	34	**
	15C	Ferrous DPD titrimetric	APHA	Dist. water	640 µg/l	47	**
					1,830	23	*
	15D	DPD colorimetric	APHA	Dist. water	660 µg/l	71	***
Fluoride	20A	Distillation-ion electrode	EPA	N.G.	0.85 mg/l	3	*
No interference	20B	SPADNS	EPA	N.G.	0.83 mg/l	24	*
interference					0.57 mg/l	41	**
No interference			APHA	N.G.	0.83 mg/l	17	*
interference					0.57 mg/l	39	**

Method Rating
* Excellent 0-25%
** Acceptable 25-50%
*** Unacceptable >50%

(continued)

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matric	True value	Total error%	
Hardness	21A	Titration	EPA	Synthetic	31 mg/l	19	*
					33	15	*
					132	6	*
					194	4	*
					417	3	*
					444	7	*
			APHA	Dist. water	680	7	*
Hydrogen ion pH	22	Electrometric measurement	EPA	Synthetic	3.5 pH units	6	*
					7.1	7	*
					7.2	5	*
					8.0	3	*
					8.0	3	*
Kjeldahl nitrogen	23A	Nesslerization	EPA	Natural water	0.2 mg/l	212	***
					0.31	166	***
	23B	Titrimetric	EPA	Natural water	4.10	52	***
					4.61	53	***
Aluminum	24A	A. A. method	EPA	Natural water	15 ug/l	430	***
					35	277	***
					500	88	***
					625	94	***
					1,004	78	***
					1,205	56	***
Boron total	34A	Colorimetric (curcumin)	EPA	Dist. water	240 ug/l	46	**
	34B	Colorimetric (carmine)	APHA	Dist. water	130 ug/l	72	***

(continued)

Method Rating
* Excellent 0-25%
** Acceptable 25-50%
*** Unacceptable >50%

TABLE A-3 (continued).

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Cadmium total	36A	A. A. method	EPA	Natural water	1.4 µg/l	349	***
					2.8	205	***
					14	177	***
					13	116	***
					71	61	***
					73	40	**
	36B	Colorimetric (Dithizone I)	APHA	N.G.	50 µg/l	55	***
Calcium total	38B	Permanganate	APHA	Dist. water	108 mg/l	10	*
	38C	EDTA titrimetric	APHA	Dist. water	108	20	*
Chromium VI total	40	Colorimetric	APHA	Dist. water	110 µg/l	112	***
Chromium total	42	A. A. method	EPA	Natural water	7.4 µg/l	249	***
					15	127	***
					74	31	***
					93	35	***
					370	61	***
					407	69	***
			APHA	N.G.	50	103	***
Copper total	46A	A. A. method	APHA	Natural water	7.5 µg/l	192	***
					12	177	***
					60	34	***
					75	60	***
					302	33	**
					332	36	**
			EPA	N.G.	1,000	26	**

(continued)

Method Rating

*	Excellent 0-25%
**	Acceptable 25-50%
***	Unacceptable >50%

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Iron total	50A	A. A. method	EPA	Natural water	10 µg/l	1,760	***
					24	716	***
					350	75	***
					438	34	***
					700	54	***
					340	43	**
	50B	Colorimetric (Phenonthraline)	APHA	Dist. water	300	64	***
Lead total	52	A. A. method	EPA	Natural water	25 µg/l	202	***
					37	145	***
					84	96	***
					101	91	***
					334	68	***
					367	73	***
Magnesium total	54A	A. A. method	APHA	N.G.	0.2 mg/l	27	**
	54B	Gravimetric	APHA	N.G.	82 mg/l	18	*
Manganese total	56A	A. A. method	EPA	Natural water	11 µg/l	584	***
					17	257	***
					84	64	***
					106	61	***
					425	34	**
					469	43	**
			APHA	N.G.	50	33	**
	56B	Colorimetric (Persulfate)	APHA	N.G.	120	53	***
		(Periodate)	APHA	N.G.	120	97	***

(continued)

Method Rating
- * Excellent 0-25%
- ** Acceptable 25-50%
- *** Unacceptable >50%

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Mercury total	58	A. A. method	EPA	River water	0.21 µg/l	329	***
					0.27	260	***
					0.51	244	***
					0.60	143	
					3.40	88	***
					4.1	62	***
					8.8	84	***
					9.6	80	***
Silica dissolved	73	Colorimetric (Molybdo-silicate)	APHA	Dist. water	5 mg/l	36	**
					15	21	*
					30	25	*
Silver total	74A	A. A. method	APHA	Dist. water	550 µg/l	46	**
	74B	Colorimetric (Dithizone)	APHA	Dist. water	150	189	***
Sodium total	76B	Gravimetric	APHA	N.G.	19.9 mg/l	23	*
	76A	Flame photometric			19.9	39	**
Vandium total	84	Colorimetric (Gallic acid)	APHA	Dist. water	6 µg/l	40	**
Zinc total	86A	A. A. method	EPA	Natural water	7 µg/l	949	***
					11	384	***
					56	111	***
					70	87	***
					231	70	***
					310	74	***
			APHA	N.G.	500	17	*

Method Rating
- * Excellent 0-25%
- ** Acceptable 25-50%
- *** Unacceptable >50%

(continued)

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Zinc total	86B	Colorimetric (Dithizone)	APHA	Dist. water	650 µg/l	62	***
Nitrate (as N)	83A	Cadium reduction	APHA	Dist. water	50 µg/l	240	***
					500	58	***
					5,000	19	*
	83B	Brucine sulfate	EPA	Natural water	160 µg/l	121	***
					190	98	***
					1,030	49	**
					1,240	38	**
			APHA	Dist. water	50 µg/l	141	***
					500	29	**
					5,000	35	**
	83C	Automated cadmium reduction	EPA	Natural water	290 µg/l	14	*
					350	71	***
					2,310	32	**
					2,480	17	*
Organic carbon total	91	Combustion infrared	EPA	Dist. water	4.9 mg/l	175	***
					107	17	*
						(Kjeldahl N)	
Organic nitrogen total	92	Kjeldahl N-Ammonia N	EPA	Natural water	0.20 mg/l	212	***
					0.31	166	***
					4.10	52	***
					4.61	53	***

(continued)

Method Rating
- *   Excellent 0-25%
- ** Acceptable 25-50%
- *** Unacceptable >50%

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
						(Ammonia N)	
Organic nitrogen total	92	Kjeldahl N-Ammonia N	EPA	Natural water	0.21 mg/l	121	***
					0.26	73	***
					1.71	29	**
					1.92	31	**
			APHA	N.G.	0.20 mg/l	208	***
					0.80	114	***
					1.50	96	***
Orthophosphate (as P)	93A	Manual ascorbic acid reduction	EPA	Natural water	0.029 mg/l	72	***
					0.038	47	**
					0.335	13	*
					0.383	14	*
			APHA	Dist. water	0.228	12	*
				River water	0.228	6	*
		Murphy and Riley		Dist. water	0.228	8	*
				River water	0.228	5	*
	93B	Automated Ascorbic acid reduction	EPA	Natural water	0.04 mg/l	113	***
					0.04	78	***
					0.29	77	***
					0.30	57	***
Total suspended	106	Glass fiber filtration	APHA	N.G.	15 mg/l	102	***
					242	30	**
					1,707	2	*

(continued)

Method Rating
- * Excellent 0-25%
- ** Acceptable 25-50%
- *** Unacceptable >50%

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Specific conductance	109	Wheatstone Bridge conductivity	EPA	N.G.	100 μmhos/cm	17	*
					106	16	*
					808	20	*
					848	23	*
					1640	18	*
					1710	19	*
Phosphorus	98	Persulfate ascorbic acid reduction	EPA	Natural water	0.110 mg/l	63	***
					0.132	89	***
					0.772	37	**
					0.882	30	**
Specific conductance	109	Wheatstone Bridge conductivity	APHA	N.G.	147.0 μmhos/cm	27	**
					228.0	20	*
					303.0	18	*
Sulfate (as SO_4)	110A	Gravimetric	EPA	N.G.	259 mg/l	11	*
	110B	Turbidimetric	EPA	N.G.	8.6 mg/l	57	***
					9.2	47	**
					110	17	*
					122	16	*
					188	10	
					199	14	*
			APHA		259 mg/l	19	*
Surfactants	113	Colorimetric	APHA	Dist. water	270 μg/LAS	40	**

Method Rating
* Excellent 0-25%
** Acceptable 25-50%
*** Unacceptable >50%

(continued)

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Bromide	11	Titrimetric	EPA	N.G.	2.8 mg/l	30	**
					5.3	31	**
					10.3	12	*
					20.3	5	*
Cyanide	17A	Distillation silver nitrate colorimetric	EPA	Industrial & domestic	0.28 mg/l	37	**
					0.62	32	**
Kjeldahl nitrogen	23A	Ion electrode method	EPA	Surface water	0.13 mg/l	14	*
					0.19	11	*
Antimony	26	A. A. method	EPA	N.G.	5.0 mg/l	7	*
					15.0	4	*
Arsenic	28	A. A. method	EPA		5 µg/l	18	*
					10	25	*
					20	26	**
Barium	30	A. A. method	EPA	Industrial domestic waste	0.4 mg/l	28	**
					2.0	26	**
Beryllium	32	A. A. method	EPA	Industrial domestic waste	0.01 mg/l	20	*
					0.05	6	*
					0.25	5	*

(continued)

Method Rating
* Excellent 0-25%
** Acceptable 25-50%
*** Unacceptable >50%

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Calcium	38	A. A. method	EPA	N.G.	9.0 mg/l	8	*
					36	4	*
Cobalt	44	A. A. method	EPA	Industrial domestic waste	0.2 mg/l	15	*
					1.0	22	*
					5.0	5	*
Magnesium	54	A. A. method	EPA	Dist. water	2.1 mg/l	10	*
					8.2	5	*
Mercury	58	A. A. (Manual cold vapor)	EPA	River water	1.35 μg/l	32	**
					3.35	19	*
					4.35	17	*
Molybdenum	60	A. A. method	EPA	Industrial domestic waste	0.3 mg/l	5	*
					1.5	7	*
					7.5	7	*
Nickel	62	A. A. method	EPA	Industrial domestic waste	0.2 mg/l	11	*
					1.0	7	*
					5.0	9	*
Potassium	67	A. A. method	EPA	Dist. water	1.6 mg/l	28	**
					6.3	18	*

Method Rating
* Excellent 0-25%
** Acceptable 25-50%
*** Unacceptable >50%

(continued)

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Selenium	71	A. A. method	EPA	N.G.	5 μg/l	24	*
					10	22	*
					15	40	**
Sodium	76	A. A. method	EPA	Dist. water	8.2 mg/l	4	*
					52	3	*
Thallium	78	A. A. method	EPA	Industrial domestic waste	0.6 mg/l	6	*
					3.0	5	*
					15.0	5	*
Tin	80	A. A. method	EPA	Industrial domestic waste	4.0 mg/l	17	*
					20.0	6	*
					60.0	3	*
Titanium	82	A. A. method	EPA	Industrial domestic waste	2.0 mg/l	10	*
					10	11	*
					50	14	*
Vanadium	84	A. A. method	EPA	Industrial domestic waste	2.0 mg/l	10	*
					10	7	*
					50	4	*
Nitrate	88	Cadmium reduction	EPA	Sewage	0.24 mg/l	3	*
					0.55	4	*
					1.04	2	*

Method Rating
* Excellent 0-25%
** Acceptable 25-50%
*** Unacceptable >50%

(continued)

TABLE A-3. (continued)

Analyte	NPDES #	Method	Source	Matrix	True value	Total error %	
Oil & grease	90	Freon extractable infrared	EPA	Sewage	14 mg/l	20	*
				Sewage	14	21	*

Method Rating
* Excellent 0-25%
** Acceptable 25-50%
*** Unacceptable >50%

Sources:
EPA - Reference G-1
APHA - Reference G-2

GLOSSARY OF STATISTICAL TERMS

Term	Symbol	Formula
Basic statistics	(where n = number of observations)	
mean	$\overline{X}$	$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
standard deviation	S	$S = \sqrt{\frac{n \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}{n(n-1)}}$
95% confidence interval		$\overline{X} \pm t(n-1,\ 0.975)\left[\frac{S}{\sqrt{n}}\right]$
coefficient of variation		$\frac{S \times 100}{\overline{X}}$
Outlier and normality tests		
T-test	T-value	$T = \frac{\lvert X_E - \overline{X}\rvert}{S}$ where: X_E = value farthest from $\overline{X}$
coefficient of kurtosis	b_2	$b_2 = n \sum_{i=1}^{n} (X_i-\overline{X})^4/[\sum_{i=1}^{n} (X_i-\overline{X})^2]^2$
coefficient of skewness	$\sqrt{b_1}$	$\sqrt{b_1} = \sqrt{n} \sum_{i=1}^{n} (X_i-\overline{X}^3/[\sum_{i=1}^{n} (X_i-\overline{X}^2]^{3/2}$

continued

GLOSSARY OF STATISTICAL TERMS (continued)

Term	Symbol	Formula
Mean-Variance Tests		
Cochran's test	C-test	$C = S^2 \text{ max.} / \sum_{i=1}^{k} S_i^2$ where: K = number of sample Si = S of replicates S max. = S of largest value
Bartlett's test	B-test	$S^2 = \frac{\Sigma \gamma i \, Si^2}{\gamma}$ where: γ=sum of degrees of freedom for each variance γ_i $C = \frac{1 + \Sigma(\frac{1}{\gamma i}) - \frac{1}{\gamma}}{3(K-1)}$ K = number of variances $B = \frac{2.303}{C} [\gamma \log S^2 - \Sigma(\gamma_i \log S_i^2)]$
(B approximates the X^2 distribution for $\gamma_i > 5$ at K-1 degrees of freedom)		
paired t-test	t-value	$d_i = \overline{X}_{m1} - \overline{X}_{m2}$ $t = \frac{\overline{d}}{S_d / \sqrt{K}}$
total error	total error = $\frac{\lvert\text{mean error}\rvert + 2S}{\text{true value}} \times 100$	

BASIC INFORMATION

State or Region ____________________ Date ____________
Contact ____________________ Interviewer ____________
Position ____________________ ____________
Phone Number ____________________
Are State/Regional labs certified? yes no

PROCEDURE

1. Is there a well defined equivalency procedure in use? yes no

If yes,
1. Is there an information packet available? yes no
2. Do methods exist for point source versus nationwide variance? How? yes no
3. Does the method,
    a. require accuracy data yes no
    b. test for outliers yes no
    c. compare variances yes no
        # of tests and type yes no
    d. If variances are different, what happens? yes no
    e. Require interlaboratory tests yes no
    f. Check the distribution yes no
4. How many samples are required, what type? ____________
5. Do you specify the number of levels tested? yes no
6. Are periodic survellance tests performed after the method has achieved preliminary approval? yes no

2. Summary of Method: ____
____
____
____

Figure A-1. Format to compile data on state or regional equivalency procedures.

3. How was the procedure developed? ______________________________

____

____

USE

1. How commonly is the method employed? ___________________________

____

____

2. How many people in your area have used the approach? ______________

____

____

3. Can we contact them? yes no
   (Who are they?) ____

____

____

____

4. Is the method applicable to chemical, bacteriological and radioactive samples? ____

____

   Has it been used for these types of analyses? ___________________

____

5. How often is the method used? ___________________________________

6. Have you had any comments by users? _____________________________

____

Figure A-1. (continued)

7. How successful has it been? ______________________________

____

8. Could you send us an example of how the procedure has been employed?

____

____

EVALUATION

1. What is your assessment of the advantages and disadvantages of the method you use? ______________________________

____

____

2. Is the method helpful to you in your job? ______________________________

3. How would you make the method better? ______________________________

____

____

COMMENTS (include indication of cooperation received)

____

____

____

____

____

Figure A-1. (continued)

Bibliography of Reviewed Literature
Grouped by Subject Matter

Bacteriology

B-1 Stuart, D. G., et. al., Membrane Filter Techniques for the Quantification of Stressed Fecal Coliforms in the Aquatic Environment, Applied and Environ. Microbiol., 34 (1): 42-46, July 1977.

B-2 Green, B. L., et. al., Two Temperature Membrane Filter Method for Enumerating Fecal Coliform Bacteria from Chlorinated Effluents, Applied and Environ. Microbiol., 33 (6): 1259-1264, June 1977

B-3 Green, B. L., et. al., Comparison of New Millipore HC with Conventional Membrane Filters for the Enumeration of Fecal Coliform Bacteria, Applied Microbiol., 31 (4): 697-699, 1975.

B-4 Schaeffer, D. J., et. al., Statistical Analysis of the Recovery of Coliform Organisms on Gelman and Millipore Membrane Filters, Applied Microbiol., pp. 605-607, October 1974.

B-5 Lin, S. D., Comparison of Membranes for Fecal Coliform Recovery in Chlorinated Effluents, Journal WPCF, 49(11): 2255-2264, November 1977.

B-6 Sladek, K. J., et. al., Statistical Interpretation of Membrane Filter Bacteria Counts, Millipore Corporation, pp. 1-14, January 10, 1975.

B-7 EPA's Research and Development Report in Wastewater Disinfection, Technology Transfer, U.S. Environmental Protection Agency, pp. 1-6, October 1977.

B-8 Lin, Shundar, Evaluation of Fecal Streptococci Tests for Chlorinated Secondary Sewage Effluents, Journal of the Environmental Engineering Division, pp. 253-267, April 1974.

B-9 Sladek, K. J., et. al., Optimum Membrane Structures for Growth of Fecal Coliform Organisms, Millipore, pp. 1-5, 1975.

B-10 Hufham, J. B., Evaluating the Membrane Fecal Coliform Test by Using E. Coli as the Indicator Organism, Applied Microbiol., 27 (4): 771-776, 1974.

Demand

D-1 Winter, J. A., Method Research Study 3, Demand Analyses, - An Evaluation of Analytical Methods for Water and Wastewater, Environmental Protection Agency, pp. 9-22, 1971.

D-2 Jirka, Andrea, M., et. al., Micro Semi-Automated Analysis of Surface and Wastewaters for Chemical Oxygen Demand, Analytical Chemistry, 47 (8): 1397-1401, July 1975.

D-3 Verstraete, W., et. al., Three Step Measurement by the Sapromat to Evaluate the BOD_5, the Mineral Imbalance and the Toxicity of Water Samples, Water Research, 8: 1077-1081, 1974.

D-4 Tifft, E. L., and Cain, B. E., Evaluation of Automated COD for Industrial and Municipal Waste, presented at the Technicon International Congress, New York, New York, pp. 11-15, June 12-14, 1972.

D-5 Tebbutt, T. H. Y., et. al., Respirometric Determination of BOD, Water Research, 10: 613-617, 1976.

D-6 Tuffey, T. J., et. al., A Critical Analysis of Warburg Respirometry for BOD Determinations of Polluted Waters, Proceedings of the 29th Industrial Waste Conference, pp. 1-8, May 7, 8, and 9, 1974.

D-7 Stephenson, Marvin E., et. al., Wet Combustion Procedure for Determination of Carbon Content in Municipal Refuse, Environmental Science and Technology, 5 (9): 799-802, September 1971.

Metals

M-1 Salotto, Vincent B., et. al., Elemental Analysis of Wastewater Sludges From 33 Wastewater Treatment Plants in the United States, paper presented at Research Symposium on Pretreatment and Ultimate Disposal of Wastewater Solids, Rutgers University, New Brunswick, N.J., pp. 8-14, May 21-22, 1974.

M-2 Lenz, Bernard L., et. al., Ion-Selective Electrode Method Compared to Standard Methods for Sodium Determination in Mill Liquors, TAPPI, pp. 2051-2055, December 1971.

M-3 Yeager, David W., et. al., Determination of Lead in Biological and Related Material by Atomic Absorption Spectrophotometry, Environmental Science & Technology, 5 (10): 1020-1022, October 1971.

M-4 Jackson, Danny R., et. al., CD Ion Activity in Soils: Evaluation of CD Ion-Selective Electrode, Environmental Science & Technology, 11 (6): 585-587, June 1977.

M-5 Winter, J. A., et. al., Interlaboratory Study of the Cold Vapor Technique for Total Mercury in Water, Water Quality Parameters, ASTM STP 573, pp. 566-580, 1975.

M-6 Raman, A., A Comparison of Two Methods for the Determination of Serum Ionic Calcium, Clin. Biochem., 5: 208-213, 1972.

M-7 Scott, Donald R., et. al., Atomic Absorption and Optical Emission Analysis of NASN Atmospheric Particulate Samples for Lead, Environmental Science & Technology, 10 (9): 877-880, September 1976.

M-8 Maines, Ian S., et. al., Determinations of Lead in Potable Waters Using Delves Cup Atomic - Absorption Spectrometer with Signal Integration, Environmental Science & Technology, 9 (6): 549-551, June 1975.

M-9 Ostling, Göran, Short Communication, - A Simplified Automated Curcumin Method for the Determination of Boron in Sea Water, Analytical Chemical Acta., 78: 507-512, 1975.

M-10 Lund, Walter, et. al., The Application of Electrodeposition Techniques to Flameless Atomic Absorption Spectrometry, Part II. "Determination of Cadmium in Sea Water", Analytical Chemical Acta., 72: 57-62, 1974.

M-11 Bassett, J., et. al., A Spectrophotometric Method for the Determination of Boron in Water by Use of Ferroin, Analyst, 99: 1-11, January 1974.

M-12 Mesman, Brick B., et. al., A Study of Two Atomic Absorption Methods for the Determination of Sub-Microgram Amounts of Arsenic and Selenium, Analytical Letters, pp. 449-459, 1975.

M-13 Jekerka, Ivan, et. al., Simultaneous Determination of Sodium, Potassium, and Ammonium Ions by Automated Direct Potentiometry, Analytical Letters, pp. 463-472, 1974.

M-14 Greenberg, Arnold E., et. al., Mercury In Water: An Evaluation of Laboratories and Methodology, Water Technology/Quality-Journal AWWA, pp. 552-554, August 1973.

M-15 Gawady, Abbers A., et. al., Automated Method for the Determination of Total and Inorganic Mercury in Water and Wastewater Samples, Analytical Chemistry, 48: 110-116, January 1976.

M-16 Tam, Kai C., Arsenic In Water by Flameless Atomic Absorption Spectrophotometry, Environmental Science & Technology, 8 (8): 734-736, August 1974.

M-17 Roldman E., et. al., A Modified Curcumin Method for Determining Trace Amounts of Boron, Water Technology/Quality-Journal AWWA, pp. 14-15, January 1975.

M-18 Elder, John F., et. al., Application of Energy-Dispersive X-ray fluorescence to Trace Metal Analysis of Natural Waters, Environmental Science & Technology, 9 (11): 1039-1042, November 1975.

M-19 Ovett, R. J., et. al., Analysis of Chromium in Natural Waters by Gas Chromatography, Environmental Science & Technology, 10 (1): 67-70, January 1976.

M-20 Robbins, D. A., et. al., Results of Trace Arsenic Analyses Performed by Various Laboratories, American Industrial Hygiene Association Journal, pp. 57-60, January 1977.

Nutrients

N-1 Winter, J. A., et. al., Method Study 2, Nutrient Analysis, Manual Methods - An Evaluation of Analytical Methods for Water and Wastewater, Environmental Protection Agency, pp. 10-46, 1970.

N-2 Johnson, David L., Simultaneous Determination of Arsenate and Phosphate in Natural Waters, Environmental Science & Technology, 5 (5): 411-414, May 1971.

N-3 Bietz, J. A., Micro-Kjeldahl Analysis by an Improved Automated Ammonia Determination Following Manual Digestion, Analytical Chemistry, 46, (11): 1617-1618, September 1974.

N-4 Yu, Kar Y., et. al., Evaluation of a Nitrate-Specific Ion Electrode, Journal WPCF, pp. 1896-1901, August 1977.

N-5 Jirka, Andrea, et. al., Ultramicro Semiautomated Method For Simultaneous Determination of Total Phosphorus and Total Kjeldahl Nitrogen in Wastewaters, Environmental Science & Technology, 10 (10): 1038-1044, October 1976.

N-6 Senn, David R., et. al., Determination of Nitrate Ion at the Part Per Billion Level in Environmental Samples with a continuous Flow Immobilized Enzyme Reaction, Analytical Chemistry, 48 (7): 954-958, June 1976.

N-7 Bremner, J. M., Use of an Ammonia Electrode for Determination of Ammonium in Kjeldahl Analysis of Soils, Comm. in Soils Science & Plant Analysis, 3 (2): 159-165, 1972.

N-8 Katsuhika, Matsunaga, A Rapid and Sensitive Method for Determination of Submicrogram Amounts of Ammonia in Fresh and Sea Water, Analytica Chimica Acta., 73 : 204-208, 1974.

N-9 Mertens, J., et. al., The Determination of Nitrate in Mineral Waters by Potentiometry and U.V. Spectrophotometry, Bull. Soc. Chim. Beig., 82 (3-4), 1973.

N-10 Osburn, Quentin W., et. al., Automated Method for Ortho-, Ortho-Plus Hydrolyzable, and Total Phosphate in Surface and Wastewaters, Environmental Science & Technology, 8 (4): 363-366, April 1974.

N-11 Thomas, Robert F., Selective Electrode Measurement of Ammonia in Water and Wastes, Environmental Science & Technology, 7 (6): 523-526, June 1973.

N-12 Langmuir, Donald, Specific Ion Electrode Determination of Nitrate in Some Freshwaters and Sewage Effluents, Environmental Science & Technology, 4 (10): 834-838, October 1970.

N-13 Heanes, D. L., Determination of Nitrate in Soil and Water by an Adaptation of an Orange 1 Method, Analyst, 100: 316-321, May 1975.

N-14 Ramirez, Munoz, The Colorimetric Determination of Phosphates in Water at Low P.P.M. Levels by Automatic Discrete - Sample Analysis, Analytical Chimica Acta., 78: 431-438, 1975.

Organics

O-1 Myers, L. H., et. al., Analytical Variability of Five Wastewater Parameters - Petroleum Refining Industry, EPA-600/2-76-234. pp. 1-67, September 1976

O-2 Carter, M. J., and Huston, M. T., Preservation of Phenolic Compounds in Wastewaters, Environmental Science & Technology, 12 (3): 309-313, March 1978.

O-3 Fountaine, J. E., New Ultraviolet Ratio Spectrophotometric System for the Determination of Trace Amounts of Phenolic Compounds, Analytical Chemistry, 46 (1): 62-66, January 1974.

O-4 Beasley, M. D., et. al., Sampling Errors in the Quantitation of Petroleum in Boston Harbor Water, Analytical Chemistry, 46 (12): 1858-1860, October 1974.

O-5 Fine, D. H., et. al., Determination of Dimethylnitrosamine in Air and Water by Thermal Energy Analysis: Validation of Analytical Procedures, Environmental Science & Technology, 11 (6): 577-580, June 1977.

O-6 Hofton M., Continuous Determination of Free Cyanide in Effluents Using Silver Ion Selective Electrode, Environmental Science & Technology, 10 (3): 277-280, March 1976.

O-7 Crisp, P. T., et. al., The Determination of Anionic Detergents with the bis (Ethylenediamine) Copper (11) Ion, Analytica Chimica Acta., 78: 391-396, 1975.

0-8 Afghan, B. K., et. al., An Improved Method for the Determination of Trace Quantities of Phenols in Natural Waters, Analytica Chimica Acta., 71: 355-366, 1974.

0-9 Gales, M. E., An Evaluation of the 3- Methyl - 2 - Benzothiazolinone - Hydrazone Method for the Determination of Phenols in Water and Waste Waters, The Analyst, 100: 841-847, December 1975.

0-10 Buelow, R. W., et. al., An Improved Method for Determining Organics by Activated Carbon Adsorption and Solvent Extraction - Pt. 11 (Test Method), JAWWA, 65: 195-199, 1973.

0-11 Buelow, R. W., et. al., An Improved Method for Determining Organics by Activated Carbon Adsorption and Solvent Extraction - Pt. 1, JAWWA, 65 : 57-72, 1973.

Physical, Mineral, Residue

P-1 Winter, J. A., et. al., An Evaluation of FWPCA Analytical Methods for Surface Waters, FWPCA Method Study 1 Mineral and Physical Analyses, pp. 1-33, June 1969.

P-2 A Preliminary Review of Analytical Methods for the Determination of Suspended Solids in Paper Industry Effluents for Compliance with EPA-NPDES Permit Terms, National Council of the Paper Industry for Air and Stream Improvement Inc., Special Report #75-01, pp. 7-10, January 1975.

P-3 Selmer-Olsen, A. R., Determination of Chloride in Aqueous Soil Extracts and Water Samples by Means of a Chloride - Selective Electrode, Analyst, 98 : 412-415, June 1973.

P-4 Finger, R. E., et. al., An Automated Spectrophotometric Suspended Solids Analysis for Activated Sludge, Journal WPCF 47 (5): 1043-1054, May 1975.

P-5 Chen, K., et. al., Solvent Extraction of Sulfur From Marine Sediment and its Determination by Gas Chromatography Environmental Science & Technology, 7 (10): 948-951, October 1973.

P-6 Hashimoto, Y., et. al., Preparation of Homogeneous Dust Sample For Comparison of Analytical Results of Atmospheric Samples, Environmental Science & Technology, 10 (8): 815-818, August 1976.

P-7 Babcock, R. H., et. al., Selective Ion Electrode System for Flouride Analysis, Journal AWWA 60: 953-961, August 1968.

P-8 MacLeod, K. E., et. al., Comparison of the Spadns - Zirconium Lake and Specific Ion Electrode Methods of Flouride Determination in Stack Emission Samples, Analytical Chemistry, 45 (7): 1272-1273, June 1973.

P-9 West, P. W., et. al., Spectrophotometric Determination of Atmospheric Flourides, Environmental Science & Technology, 4 (6): 487-491, June, 1970.

P-10 Open, B. R., et. al., A Comparative Study of Wet Chemical and Instrumental Methods for Sulfate Determination, presented at ACS 173 RR National Meeting, pp. 117-120, March 1977.

P-11 Devine, R. F., et. al., Interference of Sulfate Ion on Spadns Colorimetric Determination of Flouride in Waste Waters, Environmental Science & Technology, 9 (7): 678-679, July 1975.

P-12 Lawicki, C. R., An Evaluation of Methods for Measuring Sulfate, presented at ACS 171st National Meeting, pp. 88-89, April 1976.

P-13 Crosby, N. T., et. al., An Evaluation of Some Methods for the Determination of Flouride in Potable Waters and Other Aqueous Solutions, Analyst, 93: 643-652, October 1968.

P-14 Armstrong, F. A. J., Spectrophotometric Determination of Sulfide in Water, ASTM STP 573, pp. 14-19, 1975.

P-15 Hathaway, L. R., et. al., Indirect Determination of $SO_4^=$ in Water by Atomic Absorption, Applied Spectroscopy, 29 (6): 518-519, 1975

P-16 Robertson, D. J., et. al., Improvements in Phenoldisulfonic Acid Method for Determination of NO_x, Environmental Science & Technology, 9, (10): 979-980, October 1975.

P-17 Wang, L. K., Evaluation of Two Methylene Blue Methods for Analyzing MBAS Concentrations in Aqueous Solutions, Journal AWWA, 67: 182-184, April 1975.

P-18 Erdmann, D. E., Automated Ion-Selective Electrode Method for Determining Flouride in Natural Waters, Environmental Science & Technology 9 (3): 252-253, March 1975.

P-19 Krijgsman, W., A Simple and Rapid Semi-Automatic Micro Method for the Determination of Chloride in Serum and Urine, Clinica Chimica Acta, pp. 575, May 1970.

Radiology

R-1 Fukai, R., et. al., Intercalibration of Methods for Radionuclide Measurements on a Seaweed Sample, Ref. Meth. for Marine Radioactivity Studies, 11, pp. 213-226, 1975.

R-2 Fukai, R., et. al., Results of Plutonium Intercalibration in Seawater and Seaweed Samples, Ibid. pp. 107-115, 1975.

R-3 Kanipe, L., Handbook for Analytical Quality Control in Radiological Laboratories, E-EP/77-4, EPA 600/7-77-088, August 1977.

Interlaboratory Studies

A-1 Green, A., et. al., Development of a System for Conducting Inter-Laboratory Tests for Water Quality and Effluent Measurements, EPA-600/4-77-031, pg. 7, June 1977.

A-2 Robertson, A., IFYGL Chemical Intercomparisons, (IFYGL), Proc. 17th Conference Great Lakes Res., International Associated Great Lakes Res., pp. 682-696, 1974.

A-3 Larkin, R., et. al., Collaborative Testing of a Gas Chromatographic Charcoal Tube Method for Seven Organic Solvents, American Industrial Hygiene Association Journal, 38: 543-553, October 1977.

A-4 1977 Annual Book of ASTM Standards, Determination of Precision and Bias of Methods of Committee D-19 on Water, Part 31, pp. 7-15.

A-5 Youden, W. J., and Steiner, E. H., Statistical Manual of the Association of Offician Analytical Chemists, Association of Official Analytical Chemists, pp. 1-87, 1975.

Statistics

S-1 Shapiro, S. S., et. al., A Comparative Study of Various Tests for Normality, American Statistical Association Journal, pp. 1343-1372, December, 1968.

S-2 Hirsch, R. F., Analysis of Variable in Analytical Chemistry, Analytical Chemistry, 49 (9): 691-700, July 1977.

S-3 Grubbs, F. E., Procedures for Detecting Outlying Observations in Samples, Technometrics, 11 (1): 1-21, February 1969.

S-4 Dixon, W. J., Processing Data for Outliers, Biometrics, pp. 74-39, March 1953.

S-5 Gumbel, E. J., et. al., Statistical Control of Sample Extremes, Columbia University, School of Engineering and Applied Science, pp. 1-20, March 1967.

S-6 Draper, N. R., Applied Regression Analysis, John Wiley & Sons, Inc., pp. 12-31, 1966.

S-7 Youden, W. J., The Sample, the Procedure, and the Laboratory, Analytical Chemistry, 32 (13): 23A-37A, December 1960.

S-8 Chazal, M. De., How to Discard Invalid Measurements, Chemical Engineering, pp. 183-184, February 1967.

S-9 Quesenberry, C. P., On Testing Normality Using Several Samples: An Analysis of Peanut Aflatoxin Data, Biometrics, 32: 753-759, December 1976.

S-10 Downton, F., Nonparametric Test for Block Experiment, Biometrika, 63 (1): 137-141, 1976.

S-11 Andersen, L.B., Nonparametric Statistics Provide Comparisons Between Distributions, Chemical Engineering, pp. 139-144, July 8, 1963.

S-12 Bradley, R. A., Science, Statistics, and Paired Comparisons, Biometrics, 32: 213-232, June 1976.

S-13 Pearson, E. S., Tables of Percentage Points of $\sqrt{b_1}$ and b_2 in Normal Samples; A Rounding Off, Biometrika, pp. 282-285.

S-14 Geary, R. C., Testing for Normality, Biometrika, 34: 209-242, 1947

S-15 Moroney, M. J., Facts From Figures, pp. 232-233, 1951.

S-16 Gayen, A. K., The Distribution of the Variance Ratio in Random Samples of any Size Drawn from Non-Normal Universes, Biometrika, 37: 236-251, 1950.

S-17 Gayen, A. K, The Distribution of Student's t in Random Samples of Any Size Drawn from Non-Normal Universes, Biometrika, 36: 353-368, 1949.

S-18 Bartlet, M. S., The Use of Transformations, Biometrics, 3: 39-51, 1947.

S-19 Cochran, W. G., Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied, Biometrics, 3: 22-38, 1947

S-20 Gayen, A. K., Significance of Difference Between the Means of Two Non-Normal Samples, Biometrika, 37: 399-408, 1950.

S-21 Pearson, E. S., Relation Between the Shape of Population Distribution and the Robustness of Four Simple Test Statistics, Biometrika 62, (2): 223-241, 1975.

S-22 Anscombe, F. J., et. al., The Examination and Analysis of Residuals, Technometrics, 5 (2): 141-159.

S-23 Concepts and Applications of Regression Analysis, IBM Data Processing Application, Technical Publications Dept., White Plains, N.Y., 1966.

S-24 Daniel, C. and Wood, F. S., Fitting Equations to Data, Wiley Interscience, John Wiley & Sons, p. 32, 1971.

S-25 Bartlett, M.S., Fitting a Straight Line When Both Variables are Subject to Error, Biometrics, p. 207, September 1949.

General

G-1 Methods for Chemical Analysis of Water and Wastes, EPA-625/6-74-300, pp. 1-298, EPA Environmental Monitoring and Support Laboratory, Cincinnati, Ohio, 1974.

G-2 Standard Methods for the Examination of Water and Wastewater, pp. 61-967, 14th Edition, APHA, AWWA, WPCF, 1975.

APPENDIX B

FEASIBILITY REPORT

A report establishing the need for using a computerized data handling system is required prior to the development of a computer program. A feasibility study report entitled "Equivalency Data Analysis and Storage System" was prepared by Richard L. Prairie and printed on August 19, 1977. This report covers the data evaluation and storage system presently used for the interim Alternative Test Procedure Equivalency Testing System. Only minor additions to this report are necessary to meet the requirements of the data evaluation system proposed in this project. A copy of the August 19, 1977 feasibility report and the recommended additions to the report are included in this section of the appendix.

I. Introduction

A. History

On May 26, 1977 Mr. Edward J. Nime, director of the Computer Services and System Division (CSSD) of the Environmental Reasearch Center, Cincinnati, on behalf of the Environmental Monitoring and Support Laboratory, (EMSL) Cincinnati, requested the Southwestern Ohio Regional Computer Center to prepare "an analysis of requirements including a definition and justification for developing" and Equivalency Data Analysis and Storage System (EDASS), "and an evaluation of alternative solutions".

Mr Nime included with his request a detailed description of the data analysis requests of the system, i.e., the EDASS statistical procedures. Subsequently, a meeting with CSSD and EMSL personnel resulted in a definition of the data base storage requirements of EDASS. This information was transmitted to SWORCC in a memorandum from Bruce Almich on June 28, 1977.

B. Project References

Project Sponsor - Environmental Monitoring and Support Laboratory-Cincinnati

Statement of Work - Task Assignment No. EPA-77-48, May 26, 1977

II. System Summary

A. End Product

The Equivalency Data Analysis and Storage System consists, as the name implies, of two modules: a program for the statistical analyses of the equivalency of two analytical methods, and a data storage and retrieval system for the results of the equivalency analyses.

B. Justification

On December, 1976, the guideline establishing test procedures for the analyses of pollutant (40 CRF 136) was amended to allow EPA to approve alternative test procedures for pollutants. Applicants for such procedures are requested to submit data to establish the equivalency of the currently approved procedures and the proposed alternative test procedure.

A typical set of data will consist of 48 analyses to establish the acceptability of a new method for one parameter measured on one type of discharge. Each application may wish to establish equivalency for up to five types of discharge. From 100 to 300 applications are anticipated per year. Each application may request certification of a method for one or more parameters; the expected average is two. Each application file will consist of approximately 500 characters of information.

Thus, each year from 200 to 600 applications will be handled by EDASS. Over a 5 year period 1000 to 3000 historical records (0.5-1.5 million characters) may accumulate, which must be accessed and summarized. A statistical analysis of some 128,000 data points is anticipated per year.

C. Timing Requirements

In order to be responsive to the amount of data anticipated for EDASS, the statistical analysis package should be accessible within 10 seconds during a working day. After data input, the program should run within 10 minutes. The data base storage and retrieval should have the same accessibility requirements and, in addition, maintain records for a 5 year period.

D. System Input Requests

Data for the statistical analysis program will be input as punched card data and stored on disc for processing. The cards themselves will serve as backup data. After processing, the sample data will not be retained by the system. Rather, the results will be saved in the on-line historical data file for access by the data storage and retrieval package.

E. Processing Requirements

The data analysis program will run without human interaction unless the data or the base 10 logarithm of the data do not have equal within-sample variances. In such a case, the program will request the operator to stratify the data and then rerun the analysis.

The statistical report will be available for printing at the completion of the data analysis. A special results file will be created for data input to the data storage and retrieval package.

F. Interactions with Other Systems or Organizations

The data analysis and data base will not be shared with any other organization or system. EMSL-Cincinnati will have total control of the data and results of data analysis.

III. Feasible Alternative Systems

A. Manual System

Due to the statistical complexity of the analysis to establish equivalency of methodology as well as the large volume of the expected data, a manual manipulation of the data is totally impractical. The use of modern computer technology is the only feasible alternative to the data analysis portion of EDASS.

The data storage and retrieval system could be maintained in a filing cabinet and manually searched and tallied. A file containing the applications and copies of correspondence will be kept. However, summary reports of the number of applications by parameter, discharge type, acceptance or rejection of the alternate method, status of the application, etc., will take longer and longer to generate as the data base grows. It will be unreasonable to generate such statistics manually in a short time if an unexpected report is requested.

B. Automated Systems

Four computer systems need to be considered for EDASS:

1. Data General 840 (EMSL)
2. PDP 11/70 (CSSD)
3. Univac 1110 (NCC)
4. IBM 370/168 (COMNET)

All four systems have the capability to perform the statistical calculations. A FORTRAN program could be written which, with very slight modifications for I/O compatability, could run on any of the machines.

Only the first two systems could provide the requested turnaround time from card input. The large scale computers at NCC and COMNET would require batch input of the card data base before a time sharing job could statistically analyze the data.

Either the Data General or PDP-11 computers could perform the statistical analysis in the requested time period.

The data storage and retrieval system could be programmed specifically for EDASS, but would lack the flexibility of a generalized data base management system. Compared to the statistical analysis portion of EDASS, the data storage and analysis program would need to be considerally more complex in order to provide the capability to generate arbitrarily defined reports. A data base management system, on the other hand, already provides such a facility.

Three of the computers have or will have data base management systems - System 2000 at NCC and COMNET and a yet to be announced system at CSSD for the PDP-11/70. On the basis of system specifications, either data base management system is judged quite adequate to meet the requirements of EDASS. However, the computers using System 2000 could produce reports only via their batch facilities, with the inherently slow turnaround of the batch mode of operation.

The PDP-11/70 at CSSD, however, would be accessible on a time-sharing mode. Additionally, the results of the data analysis programs could be saved in a file for direct input into the data base management system on the same computer.

IV. Cost/Benefit Analysis of Alternatives

Since, based on performance characteristics only one feasible alternative has been identified, no detailed cost benefit analysis has been performed. However, in terms of direct cost to EMSL which includes connect time, the PDP-11/70 has been benchmarked to be at least a factor of 10 cheaper than NCC and Optimum Systems Incorporated, (OSI) which has been replaced by COMNET. Early experience with COMNET suggest that the overall costs on that system are similar to those at OSI. Thus, if NCC or COMNET had been a feasible alternative, their cost would have been at least 10 times greater than the PDP-11/70.

V. Recommendations

EDASS is recommended to be implemented on the PDP-11/70 at CSSD, Cincinnati. Of the possible computer systems only the PDP-11/70 gives satisfactory response time. In addition, the cost of the PDP-11/70 is at least an order of magnitude lower than other computers with data base management systems available to EMSL-Cincinnati.

VI. Impact of Recommended System

The data analysis portion of EDASS is a modest programming task in both size and execution. It should have no noticeable impact on the PDP-11. The data storage files could grow at a rate of 100,000 to 300,000 characters per year. By the time the data base grows to a size which might have some impact on the disc storage of the PDP-11, increased on-line storage should be available.

Addenda to Feasibility Report

1. Addendum to Section 1. (A.) History:
An addendum was made to section 11,C., "Timing Requirements", in March 1978 to include the requirements for the statistical evaluation procedure being recommended as a result of work on Contract No. 68-03-2595, "Development and Implementation of Alternate Procedures in Water and Waste Monitoring".

2. Addendum to Section 11. (D.) System Input Requests:
Timing Requirements - The data storage and retrieval system will be designed so that the search capability will have 10 second turn around time (CRT mode). The statistical program will be batch processed and the turn-around time will vary with computer load. Batch turn around time may therefore vary from one minute to daily. Extended historical data will be stored on tape, cards and xerox copies of the output.

APPENDIX C

COMPUTER PROGRAM SPECIFICATION REPORT

INTRODUCTION

Appendix C provides an overview of the equivalency data analysis and storage system. The discussion is divided into three main sections:

1. Overview of the system.
2. Statistical program.
3. Test plan.

The section containing the overview of the system discusses the INFORM-II data retrieval system and the relationship of this data retrieval system to the statistical analysis program. This relationship is graphically displayed using a system flow chart. System requirements, such as timing requirements and system flexibility will also be discussed. The section on the statistical program presents a flow chart, equations, decision points, definition of special terminology and listing of instructions in sufficient detail to permit direct programming. It is emphasized that an actual program which has been coded and debugged is not provided, only generalized program logic with examples of input, calculations and output is presented. The values from statistical tests included with the sample data have been hand calculated to provide a basis for testing the program once developed. The last section discusses a few ideas on testing and debugging the statistical program once it is developed and up on the system.

OVERVIEW OF THE SYSTEM

Figure C-1 graphically outlines the equivalency data analysis and storage system. The system has three basic components:

1. INFORM-II data retrieval system.
2. The statistical program.
3. Data storage.

The data handling and storage requirements of the system will be discussed. Some other system related requirements will also be discussed. However, those requirements that are specific to the computer system (PDP-II) or beyond the scope of this preliminary system plan will be left to the systems analyst who will implement the actual system and programs. Those requirements which will not be discussed in this document include the following:

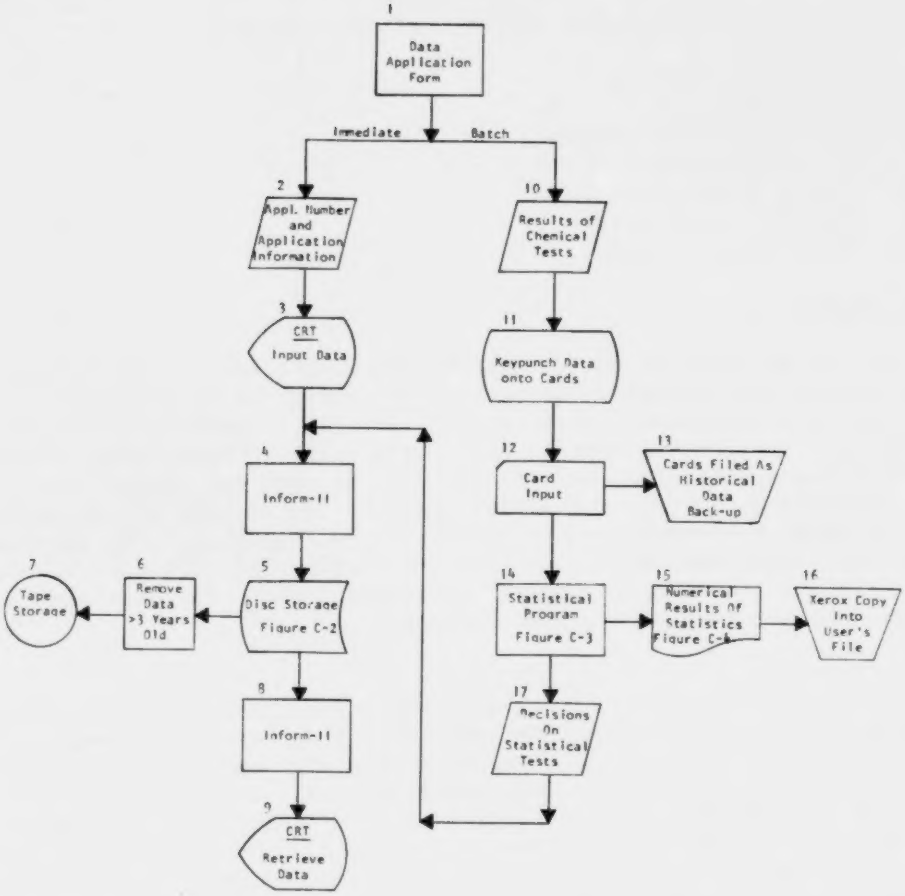

1
Data Application Form
Immediate
Batch
2
Appl. Number and Application Information
3
CRT
Input Data
4
Inform-II
5
Disc Storage
Figure C-2
6
Remove Data >3 Years Old
7
Tape Storage
8
Inform-II
9
CRT
Retrieve Data
10
Results of Chemical Tests
11
Keypunch Data onto Cards
12
Card Input
13
Cards Filed As Historical Data Back-up
14
Statistical Program
Figure C-3
15
Numerical Results Of Statistics
Figure C-4
16
Xerox Copy Into User's File
17
Decisions On Statistical Tests

Figure C-1. Overview of the system.

1. Description of the format needed to access the system.
2. Titles and tags.
3. Type and amount of security provisions.
4. Operation of the system itself.
5. Back-up recovery system.
6. Monitoring of costs.
7. Provisions to check point/restart.
8. Communications networks and technical requirements.
9. Potential interfaces with other EPA computer programs.

Topics which will be discussed in this section include:

1. System components.
    a. INFORM-II.
    b. Statistical program.
    c. Data storage.
2. Timing requirements.
3. Flexibility of system.
4. Evaluations of results.

System Components

Figure C-1 is an overview of the equivalency data analysis and storage system showing the relationships between the three system components. This system begins at location I (Figure C-1) with the information obtained from the data application form in the user's packet. Background information concerning the application, such as application name and number, are processed immediately through the INFORM system. The results of the chemical tests are batch processed through the statistical program. The application information, input data and results of the statistical tests are stored on several different devices in several different forms. The following is a detailed discussion of each system component.

INFORM System--

INFORM-II is a conversational type data retrieval system. Through the use of simple commands, this system has the ability to extract information concerning an entire application, a specific parameter for the entire data base or combinations of parameters. The operator could, therefore, extract all the information for the example Central Chemical's application (number D78-0012), all the applications during the month of December or all the limited variance applications during the month of December. A command might take the form "LIST APPLICATION, DECEMBER", to obtain all applications which were applied for during the month of December.

Location numbers 2 through 9 in Figure C-1 apply directly to the INFORM system, and contain the following components:

2. Application information - This information is obtained directly from the data application form. This background information includes (Figure C-2):

   a. Application number - an alpha-numeric designation, of the form ANN-NNNN. The letter (A) can be either D or N, designating NPDES or NIPDWR (drinking water) regulation. The two numeric symbols before the dash indicate the year of the application. The four numeric symbols following the dash are a sequential numbering system. Each application is given a unique application number.
   b. File number - an alpha-numeric designation of the form NNNNA.NN. This is an office filing system reference number, and is not unique to the application.
   c. Applicant - name and address of the person or company making the application.
   d. Location - the state and the EPA region involved.
   e. Magnitude of request - limited or nationwide.
   f. Applicable legislation - NPDES or NIPDWR.
   g. Contaminant - constituent being tested.
   h. Proposed method - method being compared to the approved method to establish equivalency.
   i. Action dates - the date application was received and technical review information.

3. Input data (CRT) - a cathode ray tube (CRT) terminal can be used to quickly input all background application data into the INFORM system.

4. Disc storage - the information processed through the INFORM system
5. will be stored on disc storage. This information includes the background application data and the statistical summary from the evaluation program (Figure C-2).

6. Tape storage - any application information greater than three years
7. old is removed from disc storage and written-out onto tape.

8. Retrieve data (CRT) - a cathode ray tube (CRT) terminal can be used
9. to quickly retrieve application information.

Statistical Program--

A detailed discussion of the statistical program itself can be found in the section headed by the same title. Figure C-1 shows the system components which are related to the statistical program. Those components are presented at locations 10 through 17 and include the following:

10. Results of chemical tests - these data are obtained from the data application form and include the following.
   a. Approved method - observations #1 and #2.
   b. Proposed method - observation #1 and #2.
   c. Sample preparation - dilution, blending, spiked.

11. Keypunch data onto cards - card input will be used for the statis-
12. tical program.

13. Historical data - cards can be filed as historical data backup.

14. Statistical program - the statistical program consists of the following sections:
   a. Calculations of means and ranges.
   b. Precision test - paired t-test on sample ranges.
   c. Accuracy test.
   d. Normality test.

TYPE OF DATA	EXAMPLE
Application number	D78-0012
File number	6922 A. 22
Applicant	Central Chemicals, Inc. Wheaton, Vermont
Location	EPA Region II, Vermont
Magnitude of request	limited
Applicable legislation	drinking water
Contaminant	total organic carbon
Proposed method	Central Chemicals variation of the combustion infra-red method of analyzing for TOC.

ACTION DATES

a. Date received	03 06 78
b. External technical review	
1. Out	03 10 78
In	03 20 78
Reviewer	Tom Doe, EMSL
Due date	03 20 78
Recommendation	Approve
2. Out	03 15 78
In	03 25 78
Reviewer	Martin Doe, MERL
Due date	03 30 78
Recommendation	Disapprove
c. EMSL technical review	
1. Due	03 30 78
2. EMSL recommendation	03 30 78
3. Approved	04 10 78
4. Returned	04 12 78

Figure C-2. Data stored on disc for the inform system.

STATISTICAL SUMMARY

Number of sources: 1

	Comparable	NON-COMPARABLE Proposed less precise than approved method	Proposed more precise than approved method
Precision	0	0	1

	Approved		Proposed	
	Mean	Range	Mean	Range
Accuracy (% recovery)	98.37	95.12 - 100.00	98.21	95.00 - 100.00

Comparability	Comparable	NON-COMPARABLE Proposed mean greater than approved mean	Approved mean greater than proposed mean
1. t-test	0	0	1

	Comparable	Non-Comparable	Regression analysis not applicable
2. regression	0	1	0

Figure C-2. (continued)

e. Data transformation.
f. Comparability tests.
   - Paired t-test on the source means.
   - Regression analysis.
   - Durbin-Watson statistic.
g. Scattergram.

It is suggested that the entire statistical program be developed using the equations and information contained within this Appendix. Parts of "canned statistical programs" such as BMD, could be substituted for sections of the statistical program presented. However, convenience of using a completely self-contained program far outweighs the small savings in core space gained with "canned programs". Even if canned programs are used, the accuracy tests and Durbin-Watson statistic would still have to be part of the main body statistical program, because these are not available on "canned programs".

15.]
16.] Statistical program output - A sample output is shown later in this section. These will be copied and placed into the applicant's file as a permanent record of the results of the statistical tests.

17. Decisions on statistical tests - Figure C-2 presents the statistical summary that is generated by the statistical program for the applicant's file on the INFORM system. These results are written out to a temporary file and later processed through the INFORM system to the applicant's permanent file on disc.

Data Storage--
Various types of data storage occur in both the INFORM-II and statistical program portions of the overall system. Data storage was discussed in the sections describing the INFORM-II system and the statistical program. However, these forms of data storage will be summarized as follows:

1. INFORM-II
   a. Current background application information is stored on disc and includes background application data and the statistical summary. This data can be readily retrieved using a cathode ray tube terminal.
   b. Application data no longer considered current will be removed from disc storage and stored on tape.
2, Statistical program.
   a. Cards used to input data into the statistical program can be stored as historical data backup.
   b. Numerical results of the statistical analyses will be printed out, xeroxed and stored in the applicant's file.

Timing Requirements

The data storage and retrieval system will be designed so that the search capability will have 10 seconds turnaround time (CRT mode). The statistical program will be batch processed and the turnaround time will vary with computer load. Batch turnaround time may, therefore, vary from one minute to daily.

Flexibility of System

As mentioned in the Feasibility Report (Appendix B), it is anticipated that from 200 to 600 applicants will be handled by the equivalency data analysis and storage system. However, by the time the data base grows to a size which might have some impact on the disc storage of the PDP-11, increased on-line storage should be available. It appears that the system is flexible enough to handle the anticipated application load.

The equivalency data array should be 120 rows by 11 columns wide. This array will handle a maximum of 10 sources with 12 samples per source, or one source with 120 samples. This should give the program enough flexibility to handle all possible data inputs. It is estimated that the program once developed, which contains this array, will require approximately 50-100K bytes of core.

Evaluation of Results

The section which follows, "Statistical Program", discusses the results of the statistical tests. However, this discussion was intended not as a guideline to critically evaluate these results but as a guide to help the programmer make logical decisions in the program flow. Refer to the user's guidelines for the critical evaluation of the results of the statistical tests.

STATISTICAL PROGRAM

A flow chart showing the conceptualized statistical computer program is presented in Figure C-3. This flow chart presents a generalized programming logic and is not a detailed FORTRAN flow chart. However, sufficient detail will be presented in this report to permit direct programming. For specific information, many flow chart symbols contain reference figure numbers which are often followed by a figure line number in parenthesis (e.g. Figure C-4 (8a)). These refer the user to specific operations, equations or output format listed on the referenced figure. Several terms are defined to improve flow chart and figure clarity:

1. Approved method - a method of analysis designated as acceptable for National Pollutant Discharge Elimination System (NPDES) or National Interim Primary Drinking Water Regulations (NIPDWR) monitoring work. Always designated as an "X" term in the formulae. Abbreviated in the flow chart as "Appr.".

2. Proposed method - a method of wastewater analysis which an applicant feels is comparable to an NPDES or NIPDWR approved method. Always designated as an "Y" term in the formulae. Abbreviated in the flow chart as "Prop.".

3. Observation - the result of one of the duplicate analyses performed on an industrial source sample.

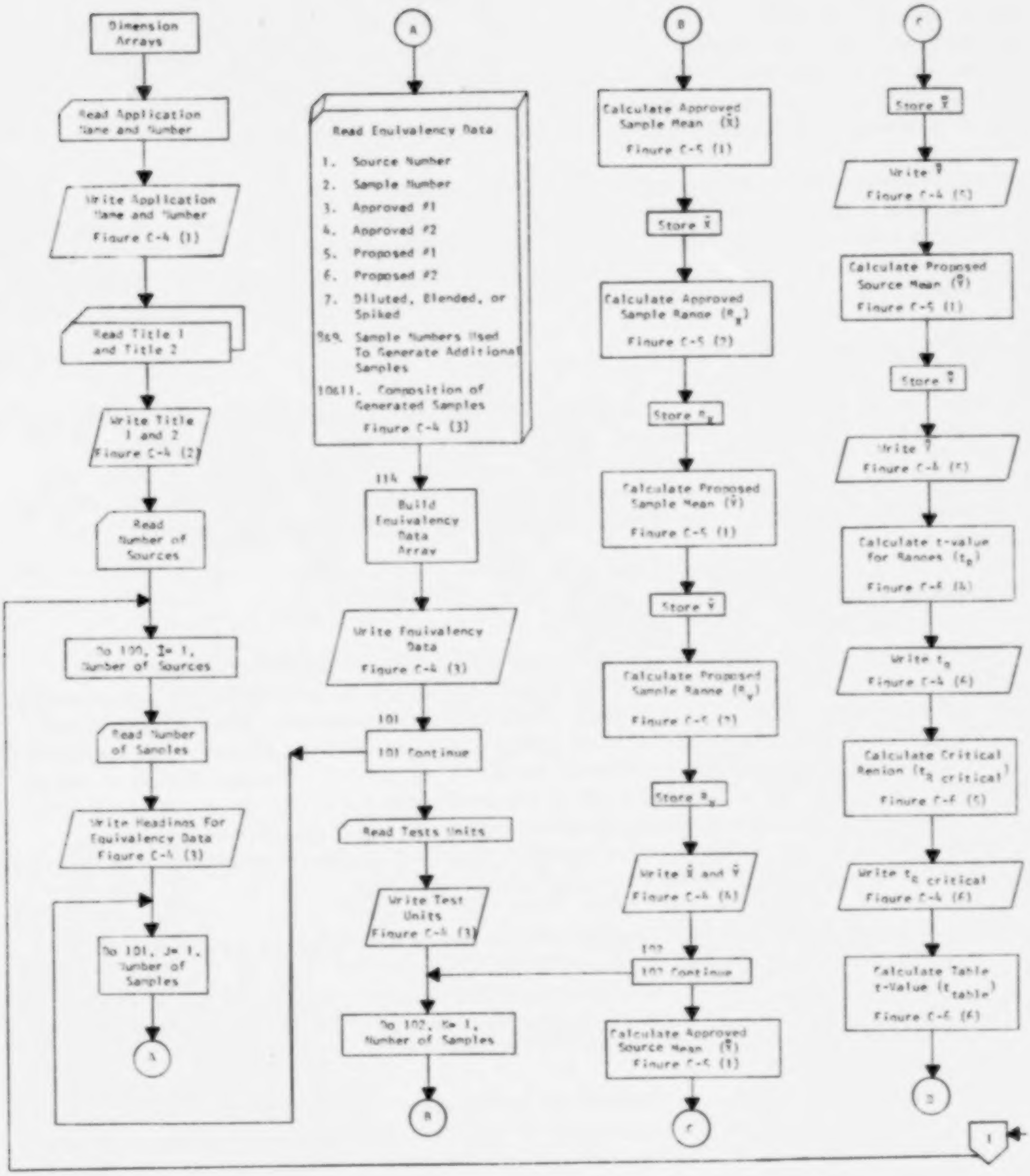

Dimension Arrays
Read Application Name and Number
Write Application Name and Number Figure C-4 (1)
Read Title 1 and Title 2
Write Title 1 and 2 Figure C-4 (2)
Read Number of Sources
Do 100, I= 1, Number of Sources
Read Number of Samples
Write Headings For Equivalency Data Figure C-4 (3)
Do 101, J= 1, Number of Samples
A
A
Read Equivalency Data
1. Source Number
2. Sample Number
3. Approved #1
4. Approved #2
5. Proposed #1
6. Proposed #2
7. Diluted, Blended, or Spiked
8&9. Sample Numbers Used To Generate Additional Samples
10&11. Composition of Generated Samples
Figure C-4 (3)
114
Build Equivalency Data Array
Write Equivalency Data Figure C-4 (3)
101
101 Continue
Read Tests Units
Write Test Units Figure C-4 (3)
Do 102, K= 1, Number of Samples
B
B
Calculate Approved Sample Mean ($\bar{X}$) Figure C-5 (1)
Store $\bar{X}$
Calculate Approved Sample Range (R_x) Figure C-5 (2)
Store R_x
Calculate Proposed Sample Mean ($\bar{Y}$) Figure C-5 (1)
Store $\bar{Y}$
Calculate Proposed Sample Range (R_y) Figure C-5 (2)
Store R_y
Write $\bar{X}$ and $\bar{Y}$ Figure C-4 (4)
102
102 Continue
Calculate Approved Source Mean ($\bar{\bar{X}}$) Figure C-5 (1)
C
C
Store $\bar{\bar{X}}$
Write $\bar{\bar{X}}$ Figure C-4 (5)
Calculate Proposed Source Mean ($\bar{\bar{Y}}$) Figure C-5 (1)
Store $\bar{\bar{Y}}$
Write $\bar{\bar{Y}}$ Figure C-4 (5)
Calculate t-value for Ranges (t_R) Figure C-6 (4)
Write t_R Figure C-4 (6)
Calculate Critical Region ($t_{R\ critical}$) Figure C-6 (5)
Write t_R critical Figure C-4 (6)
Calculate Table t-Value (t_{table}) Figure C-6 (6)
D
1

Figure C-3. Flowchart of statistical program.

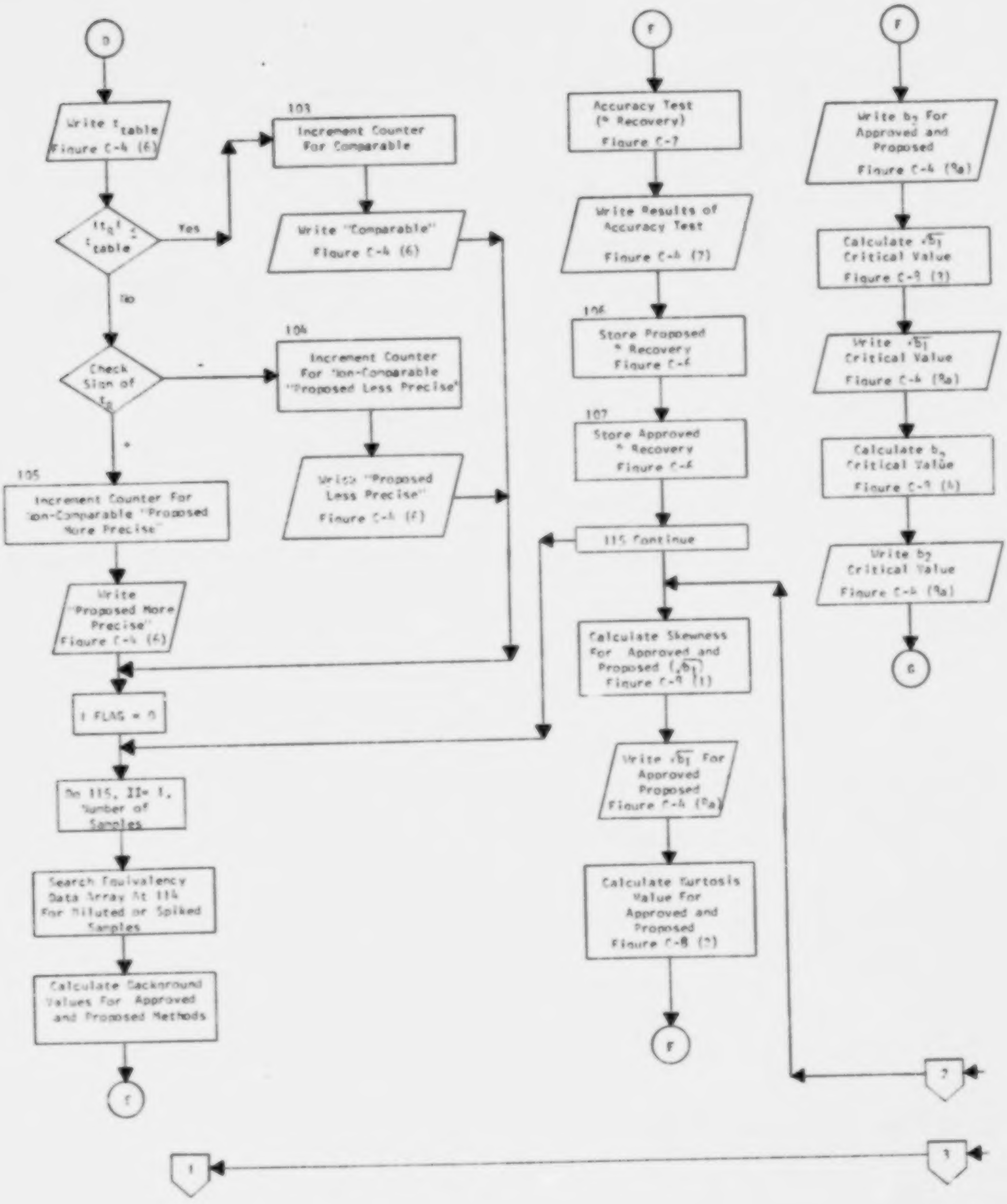

D
Write t_{table} Figure C-4 (6)
$|t_0| \leq t_{table}$
Yes
No
103
Increment Counter For Comparable
Write "Comparable" Figure C-4 (6)
Check Sign of t_0
-
+
104
Increment Counter For Non-Comparable "Proposed Less Precise"
Write "Proposed Less Precise" Figure C-4 (6)
105
Increment Counter For Non-Comparable "Proposed More Precise"
Write "Proposed More Precise" Figure C-4 (6)
I FLAG = 0
Do 115, II= 1, Number of Samples
Search Equivalency Data Array At 114 For Diluted or Spiked Samples
Calculate Background Values For Approved and Proposed Methods
E
E
Accuracy Test (% Recovery) Figure C-7
Write Results of Accuracy Test Figure C-4 (7)
106
Store Proposed % Recovery Figure C-6
107
Store Approved % Recovery Figure C-6
115 Continue
Calculate Skewness For Approved and Proposed ($\sqrt{b_1}$) Figure C-9 (1)
Write $\sqrt{b_1}$ For Approved Proposed Figure C-4 (9a)
Calculate Kurtosis Value For Approved and Proposed Figure C-8 (2)
F
F
Write b_2 For Approved and Proposed Figure C-4 (9a)
Calculate $\sqrt{b_1}$ Critical Value Figure C-9 (3)
Write $\sqrt{b_1}$ Critical Value Figure C-4 (9a)
Calculate b_2 Critical Value Figure C-9 (4)
Write b_2 Critical Value Figure C-4 (9a)
G
2
1
1

Figure C-3. continued.

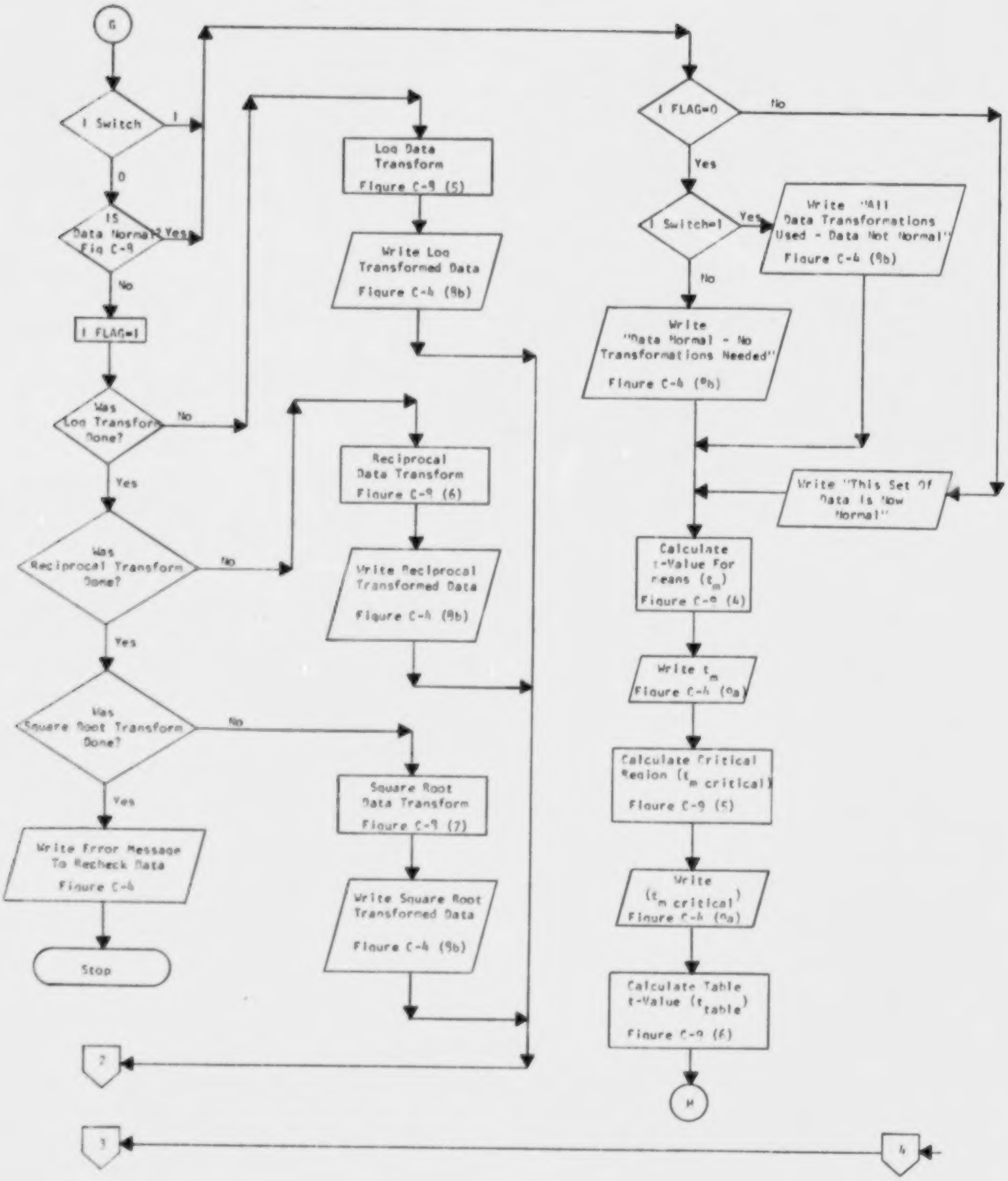

G
I Switch
1
0
IS
Data Normal?
Fig C-9
Yes
No
I FLAG=1
Was
Log Transform
Done?
No
Yes
Was
Reciprocal Transform
Done?
No
Yes
Was
Square Root Transform
Done?
No
Yes
Write Error Message
To Recheck Data
Figure C-4
Stop
Log Data
Transform
Figure C-9 (5)
Write Log
Transformed Data
Figure C-4 (9b)
Reciprocal
Data Transform
Figure C-9 (6)
Write Reciprocal
Transformed Data
Figure C-4 (9b)
Square Root
Data Transform
Figure C-9 (7)
Write Square Root
Transformed Data
Figure C-4 (9b)
2
I FLAG=0
No
Yes
I Switch=1
Yes
Write "All
Data Transformations
Used - Data Not Normal"
Figure C-4 (9b)
No
Write
"Data Normal - No
Transformations Needed"
Figure C-4 (9b)
Write "This Set Of
Data Is Now
Normal"
Calculate
t-Value For
means (t_m)
Figure C-9 (4)
Write t_m
Figure C-4 (9a)
Calculate Critical
Region (t_m critical)
Figure C-9 (5)
Write
(t_m critical)
Figure C-4 (9a)
Calculate Table
t-Value (t_{table})
Figure C-9 (6)
H
3
4

Figure C-3. continued.

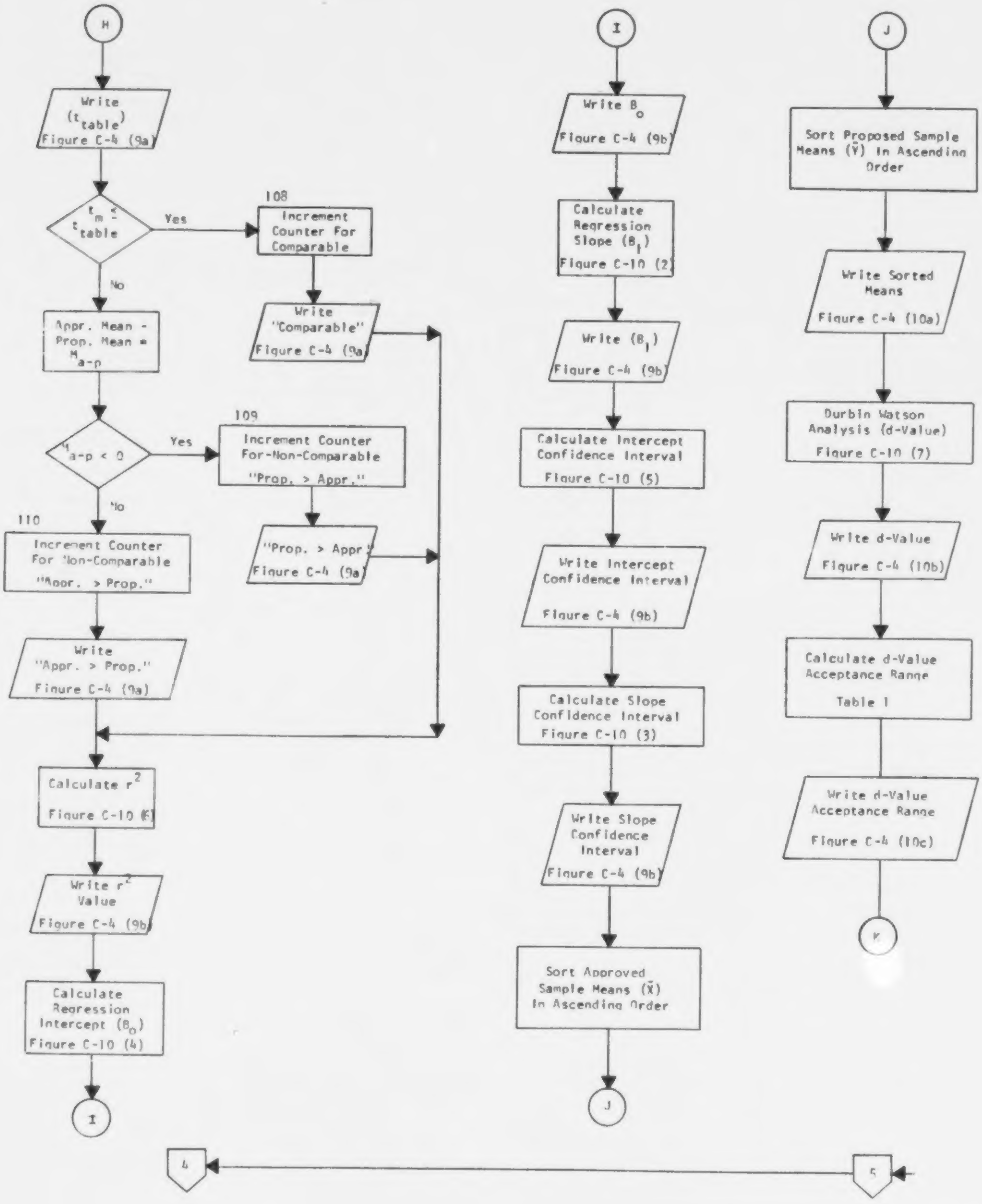

H
Write (t_{table}) Figure C-4 (9a)
$t_m \le t_{table}$
Yes
No
108
Increment Counter For Comparable
Write "Comparable" Figure C-4 (9a)
Appr. Mean - Prop. Mean = M_{a-p}
$M_{a-p} < 0$
Yes
No
109
Increment Counter For-Non-Comparable "Prop. > Appr."
"Prop. > Appr." Figure C-4 (9a)
110
Increment Counter For Non-Comparable "Appr. > Prop."
Write "Appr. > Prop." Figure C-4 (9a)
Calculate r^2 Figure C-10 (6)
Write r^2 Value Figure C-4 (9b)
Calculate Regression Intercept (B_o) Figure C-10 (4)
I
I
Write B_o Figure C-4 (9b)
Calculate Regression Slope (B_1) Figure C-10 (2)
Write (B_1) Figure C-4 (9b)
Calculate Intercept Confidence Interval Figure C-10 (5)
Write Intercept Confidence Interval Figure C-4 (9b)
Calculate Slope Confidence Interval Figure C-10 (3)
Write Slope Confidence Interval Figure C-4 (9b)
Sort Approved Sample Means ($\bar{X}$) In Ascending Order
J
J
Sort Proposed Sample Means ($\bar{Y}$) In Ascending Order
Write Sorted Means Figure C-4 (10a)
Durbin Watson Analysis (d-Value) Figure C-10 (7)
Write d-Value Figure C-4 (10b)
Calculate d-Value Acceptance Range Table 1
Write d-Value Acceptance Range Figure C-4 (10c)
K
4
5

Figure C-3. continued.

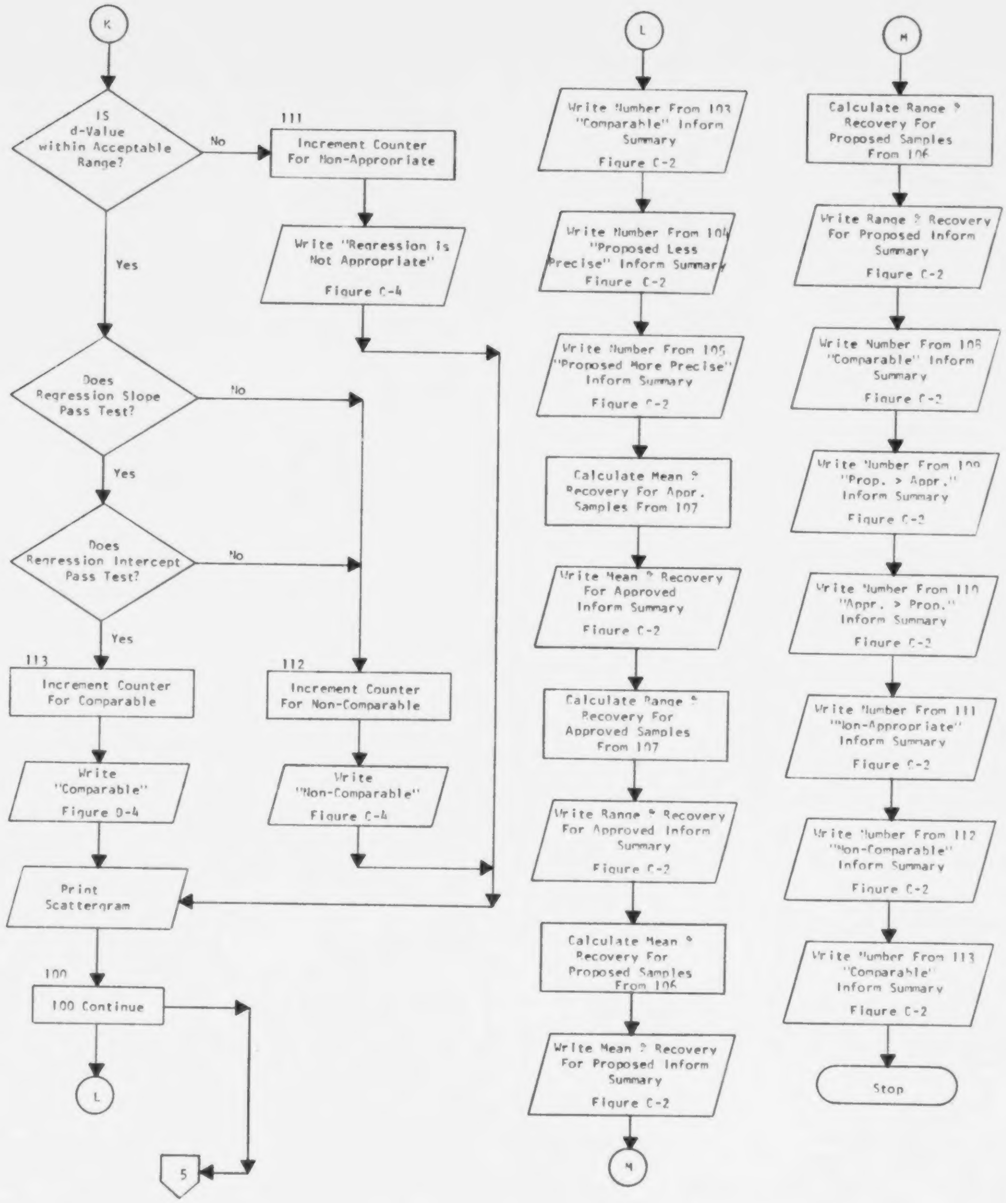

K
IS d-Value within Acceptable Range?
No
Yes
111
Increment Counter For Non-Appropriate
Write "Regression Is Not Appropriate" Figure C-4
Does Regression Slope Pass Test?
No
Yes
Does Regression Intercept Pass Test?
No
Yes
113
Increment Counter For Comparable
112
Increment Counter For Non-Comparable
Write "Comparable" Figure D-4
Write "Non-Comparable" Figure C-4
Print Scattergram
100
100 Continue
L
5
L
Write Number From 103 "Comparable" Inform Summary Figure C-2
Write Number From 104 "Proposed Less Precise" Inform Summary Figure C-2
Write Number From 105 "Proposed More Precise" Inform Summary Figure C-2
Calculate Mean % Recovery For Appr. Samples From 107
Write Mean % Recovery For Approved Inform Summary Figure C-2
Calculate Range % Recovery For Approved Samples From 107
Write Range % Recovery For Approved Inform Summary Figure C-2
Calculate Mean % Recovery For Proposed Samples From 106
Write Mean % Recovery For Proposed Inform Summary Figure C-2
M
M
Calculate Range % Recovery For Proposed Samples From 106
Write Range % Recovery For Proposed Inform Summary Figure C-2
Write Number From 108 "Comparable" Inform Summary Figure C-2
Write Number From 109 "Prop. > Appr." Inform Summary Figure C-2
Write Number From 110 "Appr. > Prop." Inform Summary Figure C-2
Write Number From 111 "Non-Appropriate" Inform Summary Figure C-2
Write Number From 112 "Non-Comparable" Inform Summary Figure C-2
Write Number From 113 "Comparable" Inform Summary Figure C-2
Stop

Figure C-3. continued.

4. Sample - one of a number (normally 1 to 6) of samples taken from a given source. The purpose of taking a number of different samples from a given source is to evaluate concentration effects upon the approved and proposed methods. Additional samples (normally 6) are obtained by diluting, spiking or blending the original samples taken from a given source.

5. Source - an industrial or municipal effluent, usually a treated wastewater for NPDES testing or a drinking water source for NIPDWR evaluation.

The program itself is divided into five major sections:

1. Input.
2. Outer program DO loop.
3. Equivalency data array DO loop.
4. Statistical evaluation.
5. Output.
    a. To the INFORM system.
    b. To the line printer.

Additional DO loops from those mentioned above may be needed to handle data manipulations and output. However, the two listed above are considered to be the major loops in the program and are necessary even in a general discussion of program logic. Discussion of the statistical program will follow the above outline.

Input

Input to the statistical program will be in card format (Figure C-3). All information to be punched onto cards will be obtained directly from the data form completed by the applicant (Figure C-2). Data input formats are presented in Table C-1. Section 3 of the output (Figure C-4) shows a listing of the input data. Card number one contains the applicant's name and application number and will appear first on the output (Figure C-4 (1)) to facilitate identification and expedite filing procedures. The second and third cards are title cards. Card number two contains:

1. Date the application was received.
2. Parameters being tested.
3. Type of variance requested (limited or nationwide variance).

The third card indicates whether the NPDES regulation or NIPDWR is applicable. The fourth card contains the number of sources which drives the outer loop of the program. The source number varies from one to n sources (normally less than 10). The fifth card contains the number of samples (normally less than or equal to twelve) which drives the inner loop. The next section of input cards contains the equivalency data itself and these can number from one to twelve cards per source. Each card within a set includes one complete data set representing a sample number (Figure C-3). Each set of equivalency data cards (source) is preceeded by a sample number card. In the example given in Figure C-4 (3), card one of the data set for source one would include all

TABLE C-1. DATA INPUT FORMAT

Variable name	Format
Application name	4A4
Application number	3A4
Title	2(20A4)
Number of sources	I4
Number of samples	I4
Source number	I4
Sample number	I4
Approved method	
observation #1	F8.0
observation #2	F8.0
Proposed method	
observation #1	F8.0
observation #2	F8.0
Diluted, blended or spiked	I4
Sample numbers used to generate additional samples	2(I4)
Concentration of additional samples	2(F8.0)
NPDES or drinking water	I4
Test units	20A4

```
                                   STATISTICAL RESULTS

                                   ****************
                                   * CENTRAL CHEM *  ------------------- 1. APPLICATION NAME AND
                                   *   D78-0012   *  -------------------    APPLICATION NUMBER
                                   ****************

                         3-6-78  TOTAL ORGANIC CARBON LIMITED VARIANCE ------------ 2. TITLE
                                 DRINKING WATER REGULATION

                              3.  COMPUTER LISTING OF INPUT DATA

                                                                          1=DILUTING
                       APPROVED      APPROVED      PROPOSED     PROPOSED  2=BLENDING
SOURCE    SAMPLE       METHOD        METHOD        METHOD       METHOD    3=SPIKING
NUMBER    NUMBER       OBSERV. #1    OBSERV. #2    OBSER. #1    OBSER. #2 OF SAMPLE   <A>    <A>    <B>     <B>
  1        100           138.00        140.00        140.00       142.00      0         0      0     0.0     0.0
  1        101            38.00         44.00         38.00        42.00      0         0      0     0.0     0.0
  1        102            63.00         58.00         59.00        62.00      0         0      0     0.0     0.0
  1        103           120.00        118.00        121.00       117.00      0         0      0     0.0     0.0
  1        104          1500.00       1450.00       1450.00      1440.00      0         0      0     0.0     0.0
  1        105           218.00        210.00        211.00       220.00      0         0      0     0.0     0.0
  1        106            17.00         22.00         17.00        21.00      1       101      0     0.5     0.0
  1        107          1045.00       1010.00       1011.00      1003.00      2       103    104     0.33    0.67
  1        108           422.00        406.00        407.00       406.00      2       102    104     0.75    0.25
  1        109           819.00        795.00        795.00       791.00      2       100    104     0.50    0.50
  1        110          2000.00       1950.00       1950.00      1940.00      3       104      0   500.0     0.0
  1        111           618.00        610.00        611.00       620.00      3       105      0   400.0     0.0

NOTES

<A>  THESE TWO COLUMNS INDICATE THE SAMPLE NUMBERS USED FOR GENERATING ADDITIONAL SAMPLES.

<B>  THESE TWO COLUMNS INDICATE THE COMPOSITION OF THE GENERATED SAMPLES.
          1.  FOR DILUTED SAMPLES, THE FRACTION INDICATES THE PORTION OF THE GENERATED SAMPLE THAT IS DILUTION WATER.
          2.  FOR BLENDED SAMPLES, THE TWO FRACTIONS INDICATE THE RELATIVE AMOUNT OF EACH SAMPLE USED FOR BLENDING.
          3.  FOR SPIKED SAMPLES, THE NUMBER INDICATES THE CONCENTRATION OF THE MATERIAL ADDED.
TEST RESULTS ARE IN mg/l.

                              4.  APPROVED AND PROPOSED SAMPLE MEANS

                    SAMPLE            APPROVED           PROPOSED
                    NUMBER            MEAN               MEAN
                     100                139.00             141.00
                     101                 41.00              40.00
                     102                 60.50              60.50
                     103                119.00             119.00
                     104               1475.00            1445.00
                     105                214.00             215.00
                     106                 19.50              19.00
                     107               1027.50            1007.00
                     108                414.00             406.50
                     109                807.00             793.00
                     110               1975.00            1945.00
                     111                614.00             615.50
                                                                                          (continued)
```

Figure C-4. Sample output from statistical program.

```
5.     APPROVED              PROPOSED
        SOURCE                SOURCE
         MEAN                  MEAN

        575.46                567.25


6.  PRECISION - RANGE COMPARISON - PAIRED "T"

    THE TABLE "T" VALUE FOR RANGE COMPARISON          =   2.19
    THE CALCULATED "T" VALUE (TR) FOR THESE DATA      =   2.56
    THE CRITICAL T-TEST RANGE IS -2.19 to +2.19

    **   THE RANGES OF THE APPROVED AND PROPOSED METHODS ARE
         NOT COMPARABLE.

         THE PROPOSED METHOD IS MORE PRECISE THAN THE APPROVED
         METHOD.


                                         7.  ACCURACY TEST

                                                                                                              TYPE
 SOURCE     SAMPLE        AMOUNT RECOVERED         BACKGROUND CONCENTRATION          % RECOVERY             1=DILUTING     SAMPLE
 NUMBER     NUMBER     APPROVED     PROPOSED       APPROVED      PROPOSED       APPROVED     PROPOSED     3=SPIKING    COMPOSITION
   1         106         19.50        19.00          41.00         40.00          95.12        95.00          1            0.50
   1         110       1975.00      1945.00        1475.00       1445.00         100.00       100.00          3          500.00
   1         111        614.00       614.00         214.00        215.50         100.00        99.63          3          400.00

                                         8.  NORMALITY TESTING

                                 8a.  SKEWNESS AND KURTOSIS CALCULATIONS
                   THE SKEWNESS FOR THE APPROVED METHOD = 1.0666
                   THE SKEWNESS FOR THE PROPOSED METHOD = 1.0675
                   THE COEFFICIENT OF KURTOSIS, APPROVED= 2.9673
                   THE COEFFICIENT OF KURTOSIS, PROPOSED= 2.9830
                   THE CRITICAL VALUE FOR SKEWNESS FOR THE APPROVED METHOD =   1.2349
                   THE CRITICAL VALUE FOR KURTOSIS FOR THE APPROVED METHOD =   4.6881
                   THE CRITICAL VALUE FOR SKEWNESS FOR THE PROPOSED METHOD =   1.2349
                   THE CRITICAL VALUE FOR KURTOSIS FOR THE PROPOSED METHOD =   4.6881

                        8b.  **THIS SET OF DATA IS NORMAL---NO TRANSFORMATION NEEDED
```

(continued)

Figure C-4. continue.

```
                    9.  COMPARABILITY ANALYSIS

       9a.  PAIRED "T" OF SAMPLE MEANS

            THE TABLE "T" VALUE FOR MEAN COMPARISON          =   2.19
            THE CALCULATED "T" VALUE (Tm) FOR THESE DATA     =   2.31
            THE CRITICAL T-TEST RANGE IS -2.19 to +2.19

            **  THE MEANS OF THE APPROVED AND PROPOSED METHODS ARE NOT
                COMPARABLE.

                THE APPROVED MEAN IS GREATER THAN THE PROPOSED MEAN



                    METHOD COMPARISON <<"R" SQUARED>>
                   THE "R SQUARE" VALUE IS,        1.0000


                    METHOD COMPARISON <<REGRESSION>>
INTERCEPT (Bo)  =          2.2593
CONFIDENCE VALUE FOR INTERCEPT  =          1.3698
THE RANGE OF THE CONFIDENCE INTERVAL IS,          3.6290  TO          0.8895


SLOPE (B1)  =              0.9818
CONFIDENCE VALUE FOR SLOPE  =              0.0041
THE RANGE OF THE CONFIDENCE INTERVAL IS,          0.9859  TO          0.9777
```

(continued)

Figure C-4. continue.

```
                    10.  DURBIN-WATSON ANALYSIS

             METHOD COMPARISON <<DURBIN-WATSON STATISTIC>>

10a.  PAIRED, SORTED SAMPLE MEANS RANKED SMALLEST TO LARGEST VALUE FOR USE IN
      THE DURBIN-WATSON STATISTIC.

             APPROVED                      PROPOSED
              19.5000                       19.0000
              41.0000                       40.0000
              60.5000                       60.5000
             119.0000                      119.0000
             139.0000                      141.0000
             214.0000                      215.5000
             414.0000                      406.5000
             614.0000                      615.5000
             807.0000                      793.0000
            1027.5000                     1007.0000
            1475.0000                     1445.0000
            1975.0000                     1945.0000


10b.  THE DURBIN-WATSON STATISTICAL VALUE IS        2.1221


10c.  THE ACCEPTABLE RANGE FOR DURBIN-WATSON IS  =  0.81   to   3.19

                    RESULTS OF REGRESSION ANALYSIS

**   APPROVED AND PROPOSED METHODS ARE NOT COMPARABLE.
```

Figure C-4. continue.

the information on that line following the source number one and sample number one hundred. Therefore, the data on each card (Figure C-3 and C-4) would include:

1. Source number.
2. Sample number.
3. Two observations for the approved method - test #1 and #2.
4.
5. Two observations for the proposed method - test #1 and #2.
6.
7. Type of sample.
    0 = tests performed on a portion of the original sample.
    1 = tests performed on a sample obtained by diluting a portion of the original sample.
    2 = tests performed on a sample obtained by blending portions of two samples.
    3 = tests performed on a sample obtained by spiking a portion of the original sample.
8. Sample numbers used to generate additional samples. In the case of a blended sample two numbers are required. In the case of a diluted or spiked sample only one number is necessary.
9.
10. Composition of generated samples. In the case of a blended sample, two numbers are necessary to indicate the relative amount of each original sample used for blending the additional sample. For example, 33% (0.33) of sample #103 and 67% (0.67) of sample #104 make-up the additional sample #107 (Figure C-4(3)). In the case of a diluted or spiked sample only one number is needed. For example, additional sample #106 (Figure C-4) was obtained by the addition of equal volumes distilled water and sample #101. Therefore, the fraction of original sample (#101) used to generate sample #106 was 0.50 (50 percent).
11.

The final card contains the units for the test results such as mg/1, ppb, %, etc. This information is written out beneath the input listing (Figure C-4(3).

Outer DO Loop

The outer DO loop 100 is controlled by the number of sources (Figure C-3). Essentially the inner part of the program which involves a series of data manipulations and statistical tests on a sample group, is executed by the loop as many times as there are sources. The number of sources is normally one for a limited variance request and several for a nationwide variance request. Except for the initial four reads and the writes to the INFORM system, all other reads, writes, data manipulations and statistical tests are done within this loop. The sample equivalency data presented in Figure C-4 which will be used to illustrate program logic, has only one source. Therefore, the outer loop is executed only once in this example.

Equivalency Data Array Loop

The equivalency data array DO loop 101 is governed by the number of samples (Figure C-3). The number of samples per source is normally less than or equal to twelve. However, for a nationwide variance the statistician may first run the equivalency data set-up as several sources with twelve or less samples per source, and then rerun the program with the equivalency data set-up as one source with many samples. For example, a nationwide variance may include five sources with twelve samples per source. The first run would have the equivalency data set-up as five sources with twelve samples per source. The second run would have the equivalency data set-up as one source consisting of sixty samples.

Within the inner loop all sample cards for a particular source are read in, building an array which contains all equivalency data. This array is also written out within the loop (Figure C-4(3)).

Statistical Evaluation

Mean and Range--
The first step in the statistical analyses is to calculate the sample mean (Figure C-5(1)) and sample range (Figure C-5(2)) for the approved and proposed method. These calculations are performed on the two observation values (Figure C-4(3)) for each of the proposed and approved methods associated with an individual sample. These values for each sample are contained in one line of the equivalency data array. Therefore, a DO loop (102) is required to perform this series of calculations (Figure C-3). Sample mean values and the corresponding sample numbers are also written out within the DO loop (Figure C-4(4)). Once the sample means have been calculated, the mean of the sample means, or source mean (Figure C-5(1)), can be computed and written out (Figure C-4(5)). All three initial calculations; sample mean, sample range and source mean, are the basis for the statistical calculations which follow and must be stored at this point until needed (Figure C-3).

Precision Test--
The test for precision is a paired t-test to statistically evaluate the comparability of the sample ranges, R_x and R_y (Figure C-5(2)). The calculated t-value (t_R) can be computed using the formulae presented in Figure C-6 (1-4). This calculated t-value (t_R) can have either a positive or negative sign. The formulae to calculate the critical region (t_R critical) and table t-value (t_{table}) are also presented in Figure C-6 (5 and 6). The table t-value, as the name implies, is commonly obtained from a table of critical values of Student's t-distribution. For purposes of programming, the positive value of the upper limit of the critical region can be used as the table t-value to which the absolute calculated t-value (t_R) can be compared if the number of samples is less than 30 (Table C-6(6)). However, if the number of samples is 31 to 100, the table t-value is 2.00, and if the number of samples is greater than 100, the table t-value is 1.96.

The flow chart (Figure C-3) shows that after the calculated t-value (t_R),

1. MEANS

$\bar{X}_i = \frac{X_i + X(i + 1)}{2}$ approved sample mean

$\bar{\bar{X}} = \frac{\Sigma\bar{X}_i}{k}$ approved source mean

$\bar{Y}_i = Y_i + \frac{Y(i + 1)}{2}$ proposed sample mean

$\bar{\bar{Y}} = \frac{\Sigma\bar{Y}_i}{k}$ proposed source mean

where:

X_i = individual sample observation, approved method

Y_i = individual sample observation, proposed method

k = number of samples

2. RANGE

$R_x = |X_i - X_{(i + 1)}|$ sample range, approved method

$R_y = |Y_i - Y(i + 1)|$ sample range, proposed method

Figure C-5. Formulate to calculate means and ranges.

1. SAMPLE RANGE CALCULATION BETWEEN APPROVED AND PROPOSED METHODS

$$D_R = R_{xi} - R_{yi}$$

where:

R_x = sample range, approved method

R_y = sample range, proposed method

2. NORMALIZE D_R

$$\bar{D}_R = \frac{\Sigma\ D_R}{k}$$

where:

k = number of samples

3. CALCULATE STANDARD DEVIATION OF THE RANGE

$$S_{D_R} = \sqrt{\frac{(\Sigma\ D^{R2}) - \{(\Sigma D_R)^2/K\}}{k-1}}$$

$$S_{\bar{D}_R} = \frac{S_{D_R}}{\sqrt{k}}$$

4. CALCULATE THE t-VALUE FOR COMPARING SOURCE RANGES

$$t_R = \frac{\bar{D}_R}{S_{D_R}}$$

for k-1 degrees of freedom

5. CALCULATE CRITICAL TEST t-VALUE

$$t_R \text{ critical} = \pm \{_e\ (3.205/n) + 0.85\}$$

where:

n = number of observation

6. CALCULATE TABLE t-VALUE

$t_{table} = + t_R$ critical if $k < 30$

$t_{table} = 2.00$ if $k = 31$ to 100

$t_{table} = 1.96$ if $k > 100$

Figure C-6. Formulae for precision test (paired t-test).

the critical region (t_R critical) and the table t-value (t_{table}) have been computed, a series of decisions are made. The first decision involves establishing the comparability of the two sample ranges, R_x and R_y. If the absolute value of the calculated t (t_R) is less than or equal to the table t-value (t_{table}), the sample ranges are comparable. If the opposite is true, the sample ranges are not comparable. If the sample ranges are comparable, no further decisions are required under the precision testing portion of the program.

If the ranges are found to be non-comparable, another decision block is entered. The calculated t-value (t_R) is then compared against the critical region ($t_R$critical) of the two tailed t-test value, with k-1 degrees of freedom at the 95% confidence level. Note that the range of the proposed method is always subtracted from the range of the approved method. Therefore, if the t_R-value lies within the lower critical region, the ranges of the proposed method are significantly larger than those of the approved method, i.e., the proposed method is not as precise as the approved method. The opposite argument applies if the t_R-value lies within the upper critical region. For purposes of programming (Figure C-3), these tests can be made by examining the sign of the calculated t-value (t_R). If the sign is negative, the proposed method is not as precise as the approved method. If the sign is positive, the opposite is true. Each time these decision steps for the precision test are encountered by the program (governed by the number sources), a counter must be incremented to keep track of the three possible outcomes (locations 103, 104 and 105) to the two decision blocks (Figure C-3). These values are needed later for the INFORM summary (Figure C-2).

The numerical output from the precision tests should be;

1. Calculated t-value (t_R).
2. Table t-value (t_{table}).
3. The critical t-test range.

Figure C-4(6) shows the calculated t-value (t_R) to be 2.56 and the table t-value (t_{table}) to be 2.19 for the data presented as an example. Because the absolute value of the calculated "t" is greater than the table t-value, the output (Figure C-4(6)) states that, "the ranges of the approved and proposed methods are not comparable". Because the sign of the calculated t-value is positive, the output (Figure C-4(6)) states that, "the proposed method is less precise than the approved method".

Accuracy Test--
The accuracy test involves calculating the % recovery from a spiked or diluted sample. Percent recovery is computed using the formulae in Figure C-7.

1. Spiked

$$\left(\frac{\text{amount recovered - background concentration}}{\text{amount added}}\right) \times 100 = \%\ \text{recovery}$$

2. Diluted

$$\left(\frac{\text{amount recovered}}{\text{Background concentration} \times \text{dilution factor}}\right) \times 100 = \%\ \text{recovery}$$

Note: For each sample involved, the calculations must be performed twice. Once for the approved method and once for the proposed method.

Figure C-7. Formulae for accuracy test.

A spiked sample involves addition of a known concentration of the constituent being tested to a portion of the original sample. The spiked sample is then tested using both the approved and proposed methods. The background concentration, which is the sample mean of the original sample is subtracted from the test results (amount recovered). The value obtained from this evaluation is then divided by the amount added and this quantity is multiplied by 100 to obtain percent recovery. To obtain a diluted sample, the original sample is diluted by a known quantity (dilution factor) of reagent grade water. For a diluted sample, the procedure to calculate percent recovery is similar. The amount recovered is divided by the product of the background concentration times the dilution factor. The resulting value is multiplied by 100 to obtain percent recovery.

For example, sample #106 is an additional sample obtained by diluting sample #101 (Figure C-4(3)). A volume of original sample was added to an equal volume of reagent grade water (i.e., a dilution factor of 50 percent). The background concentrations are calculated from sample #101 as follows.

Approved method: $$\begin{array}{r} 38.00 \\ +44.00 \\ \hline 82.00 \end{array} \div 2 = 41.00$$

Proposed method: $$\begin{array}{r} 38.00 \\ +42.00 \\ \hline 80.00 \end{array} \div 2 = 40.00$$

The amount recovered using the approved and proposed method are calculated as follows:

Approved method: $$\begin{array}{r} 17.00 \\ +22.00 \\ \hline 39.00 \end{array} \div 2 = 19.50$$

Proposed method: $$\begin{array}{r} 17.00 \\ +21.00 \\ \hline 38.00 \end{array} \div 2 = 19.00$$

The percent records can now be calculated as follows:

Approved method: $$\left(\frac{19.5}{41.0\times0.5}\right) \times 100 = 95.12\%$$

Proposed method: $$\left(\frac{19.0}{40.0\times0.5}\right) \times 100 = 95.00\%$$

The above results are presented in Figure C-4 (3).

In Figure C-4 (3), sample #110 is an additional sample obtained by spiking sample #104 with a known quantity of TOC (500 mg/1). The background concentrations are calculated from sample #104 as follows.

Approved method: $$\begin{array}{r} 1500.00 \\ +1450.00 \\ \hline 2950.00 \end{array} \div 2 = 1475.00$$

Proposed method: $$\begin{array}{r} 1450.00 \\ +1440.00 \\ \hline 2890.00 \end{array} \div 2 = 1445.00$$

The amount recovered using the approved and proposed method are calculated as follows:

Approved method: $$\begin{array}{r} 2000.00 \\ +1950.00 \\ \hline 3950.00 \end{array} \div 2 = 1975.00$$

Proposed method: $$\begin{array}{r} 1950.00 \\ +1940.00 \\ \hline 3890.00 \end{array} \div 2 = 1945.00$$

The percent recovered can now be calculated as follows:

Approved method: $$\left(\frac{1975.00-1475.00}{500}\right) \times 100 = 100\%$$

Proposed method: $$\left(\frac{1945.00-1445.00}{500}\right) \times 100 = 100\%$$

The above results are presented in Figure C-4 (7).

The accuracy test is performed within DO loop 115 which it controlled by the number of samples. The loop is needed to search the equivalency data array, location 114 on Figure C-3, for diluted (1) and spiked (3) samples (column 9 in Figure C-4 (3)). Once a diluted or spiked sample is located in the equivalency data array, the original sample must be located using column 8 labeled "<A>" in Figure C-4 (3). Once the original sample is located, the value necessary to calculate the background concentration for the approved and proposed methods can be obtained. The dilution factor or concentration of the spike is obtained from column 10 labeled "<B>" in Figure C-4(3) and brought into the loop. The additional sample test results for the approved method (columns 3 and 4 in Figure C-4(3)) and the proposed method (columns 5 and 6 in Figure C-4(3)) are used to compute the amount recovered. Once these calculations and operations have been accomplished, the percent recovery can be calculated.

The values used to compute percent recovery and the percent recovery values themselves are written out in Figure C-4(7). The percent recovery for the proposed method (location 106) and for the approved method (location 107) are stored (Figure C-3). After all sources have been read in and all samples within these sources tested for accuracy, a range and mean percent recovery for the approved method and proposed method can be calculated and written to the INFORM statistical summary (Figure C-2). The range and mean percent recovery are calculated outside the outer program DO loop 100.

Normality Test and Data Transformations--

The next portion of the program checks data normality before proceeding with the comparability tests. If the data are not normal, three types of data transformations are performed:

1. Log transformation.
2. Reciprocal transformation.
3. Square root transformation.

The data are checked for normality after each transformation. The formulae for normality tests and data transformations are presented in Figure C-8.

The test for normality entails a comparison of the calculated coefficient of skewness, $\sqrt{b_1}$ (Figure C-8(1)) and the coefficient of kurtosis, b_2 (Figure C-8(2)), to their calculated critical values for each method (Figure C-8 (3 and 4)). A method is considered normal only when both the $\sqrt{b_1}$ and the b_2 values are less than their critical values. There are five possible consequences of applying this statistic, and these are as follows:

1. Data from both methods are shown to be normally distributed, no transformations are necessary.
2. Data from both methods are non-normal and cannot be made to fit a Gaussian distribution by the three transformations presented in Figure C-8.
3. Data from one method is normally distributed, the other is not. The three transformations given do not produce normally distributed data on both methods.
4. One method produced normally distributed data, the other did not. One of the three given transformations makes them both normally distributed.
5. Both methods produced non-normal data. Data from both are normally distributed after transformation.

Statistically, only 1, 2 and 5 could occur to data from comparable methods. However, the statistics can be analyzed anyway if the data analyst so desires. For purposes of programming, 1, 4 and 5 will cause the program flow

1. CALCULATE COEFFICIENT OF SKEWNESS

$$\sqrt{b_1} = \sqrt{k}\ [\Sigma(\bar{X}_i - \bar{\bar{X}})^3 / <\Sigma(\bar{X}_i - \bar{\bar{X}})^2>^{1.5}]$$

where:
$\underline{k}$ = number of samples
$\bar{X}$ = approved sample mean
$\bar{\bar{X}}$ = approved source mean

2. CALCULATE COEFFICIENT OF KURTOSIS

$$b_2 = k[\Sigma(\bar{X}_i - \bar{\bar{X}})^4 / <\Sigma(\bar{X} - \bar{\bar{X}})^2>^2]$$

3. CALCULATE CRITICAL VALUE FOR SKEWNESS

$\sqrt{b_1}$ critical value $= 1.4244e^{-0.0119k}$ for k samples

4. CALCULATE CRITICAL VALUE FOR KURTOSIS

b_2 critical value $= 0.839 + (1.549 \ln k)$ for k=5 to 20

b_2 critical value $= 6.4868 - (0.437 \ln k)$ for k=21 to 100

5. LOG DATA TRANSFORMATION

$$\log X_i = \log_{10}(X_i + 10^6)$$

where:
X = approved observations

6. RECIPROCAL DATA TRANSFORMATION

$$\text{reciprocal } X_i = 1/(X_i + 10^6)$$

7. SQUARE ROOT DATA TRANSFORMATIONS

$$\text{square root } X_i = \sqrt{(X_i + 10^6)}$$

Figure C-8. Distribution tests and data transformations.

Note: The above equations are set up for x observations (approved observations). However, the calculations must also be performed for y observations (proposed observations). The value 10^6 has been added to the transformation formulae to eliminate the possibility of handling zeros in the transform.

Figure C-8. (continued).

to continue to the comparability tests while 2 and 3 will cause the program to print the error message to "recheck the data".

The program section on normality testing and data transformations begins by calculating skewness and kurtosis for the approved and proposed methods, and by computing the critical value for skewness and kurtosis (Figure C-3). All the values from these calculations are written out (Figure C-4(8a)). The data is then tested for normality as previously discussed. If the data distribution is normal without any transformations, a caption to this affect is written out (Figure C-4(8b)) and the program flow continues to the comparability tests. A flag (IFLAG) checks program flow for data transformations so that the correct captions are written out.

If the data distribution is non-normal, the program flow proceeds through the data transformation section. After each data transformation, a check for normality is encountered. If one of the data transformations produces data which are normally distributed, the transformed data are written out with an appropriate caption. Figure C-4(8b) does not show this situation because the data tested normal without transformations. However, line 8b of the sample output (Figure C-4) is where the transformed data and an appropriate caption would appear. Once data from one of the transforms tests normal, the program flow continues to the comparability tests.

If all the data transformations are tried and none produce a distribution which is normal, an error message to "recheck the data" is printed out and program flow is halted. The statistician checks the data for errors. If an error is found, the data is simply resubmitted for processing by the statistical program. If no errors are found, the statistician may force program flow to bypass the data transformation section, routing program flow directly to the comparability tests. This is accomplished by setting a program switch (ISWITCH) to signal non-normal data distribution. The switch value, one or zero, should be entered from one of the input cards inside the outer program loop. Data for each source must be handled separately by program flow. When ISWITCH is set to one, a caption indicating that all data transformations were tried and the data distribution was found to be non-normal is printed in section 8b of the output (Figure C-4).

Comparability Test--

Two statistical tests are used to evaluate comparability of the approved and proposed methods. These are a paired t-test and regression analysis which includes the Durbin-Watson Statistic. The formulae for the comparability tests are presented on Figure C-9 (paired t-test) and Figure C-10 (regression analysis).

A paired t-test on method means is used to evaluate method comparability. This paired t-test used to evaluate method mean equality in most ways parallels the paired t-test used to evaluate the method ranges (precision test). The results of the test which compares the calculated t-value (t_m) to the table t-value (t_{table}) can be interpreted in the same manner as in the precision test. Also, the table t-value can be calculated in the same manner as in the precision test (Figure C-9).

1. SAMPLE MEAN CALCULATION BETWEEN APPROVED AND PROPOSED METHODS

$$Dm = \bar{X}_i - \bar{Y}_1$$

where:

$\bar{X}$ = approved sample mean

$\bar{Y}$ = proposed sample mean

2. NORMALIZE Dm

$$\bar{Dm} = \frac{\Sigma\ Dm}{k}$$

where:

k = number of samples

3. CALCULATE STANDARD DEVIATION OF THE MEAN

$$S_{Dm} = \sqrt{\frac{\Sigma Dm^2 - \ \{(\Sigma Dm)^2/k\}}{k-1}}$$

$$S_{\bar{Dm}} = \frac{S_{\bar{Dm}}}{\sqrt{k}}$$

4. CALCULATE THE t-VALUE FOR COMPARING SOURCE MEANS

$$t_m = \frac{\bar{Dm}}{\sqrt{k}} \qquad \text{for k-1 degrees of freedom}$$

5. CALCULATE CRITICAL TEST t-VALUE

$$t_m \text{ critical} = \pm\ \{e^{(3.205/n)} + 0.85\}$$

where:

n = number of observations

6. CALCULATE TABLE t-VALUE

$t_{table} = t_m$ critical if $K < 30$

$t_{table} = 2.00$ if $k = 31$ to 100

$t_{table} = 1.96$ if $k > 100$

Figure C-9. **Formulae for comparability test (paired t - test).**

1. MODEL EQUATION IS OF THE FORM;

$$\bar{Y} = B_o + B_1\bar{X}$$

where;

$\bar{Y}$ = dependent variable, proposed sample mean

$\bar{X}$ = independent variable, approved sample mean

B_o = intercept

B_1 = slope

2. CALCULATE SLOPE OF THE REGRESSION EQUATION

$$S_{xx} = \Sigma\bar{X}^2 - [(\Sigma\bar{X})^2/k)]$$

$$B_1 = \frac{(\Sigma\bar{X}\bar{Y}) - [<(\Sigma\bar{X})\ (\Sigma\bar{Y})>/k]}{S_{xx}} \qquad \text{slope}$$

where;
k = number of samples

3. CALCULATE B_1 CONFIDENCE INTERVALS

$$S_{yy} = (\Sigma\bar{Y}^2) - [(\Sigma\bar{Y})^2 / k]$$

$$S_y = \left[\frac{S_{yy} - (B_1^2 * S_{xx})}{k-2}\right]^{1/2} \qquad \text{variance due to regression}$$

$$B_1 \text{ confidence intervals } = \pm \frac{t * S_y}{\sqrt{S_{xx}}}$$

where:
t = table t or calculated t value, with k-2 degrees of freedom

Figure C-10. Formulae for regression analysis.

4. CALCULATE INTERCEPT OF THE REGRESSION EQUATION

$$B_o = \bar{\bar{Y}} - (B_1 \bar{\bar{X}})$$

where;

$\bar{\bar{Y}}$ = proposed source mean

$\bar{\bar{X}}$ = approved source mean

5. CALCULATE B_o CONFIDENCE INTERVALS

$$B_o \text{ confidence intervals} = \pm \; t * S_y \left(\frac{1}{k} + \frac{\bar{\bar{X}}^2}{S_{xx}} \right)$$

where;

t = table t-value, with k-2 degrees of freedom

6. CALCULATE COEFFICIENT OF DETERMINATION

$$r^2 = \frac{[\Sigma\bar{X}\bar{Y} - <((\Sigma\bar{X}) \; (\Sigma\bar{Y}))/k]^2}{\left[\Sigma\bar{X}^2 - \left(\frac{(\Sigma\bar{X})^2}{k}\right)\right]\left[\Sigma\bar{Y}^2 - \left(\frac{(\Sigma\bar{Y})^2}{k}\right)\right]} \quad \text{coefficient of determination}$$

7. CALCULATE DURBIN-WATSON STATISTIC

$$\hat{\bar{Y}}_i = B_o + B_1 \bar{X}_i \quad \text{predicted } \bar{Y}_i$$

$$\bar{Y}_i \text{ observed} - \hat{\bar{Y}}_i \text{ predicted} = e_i \quad \text{residual (error)}$$

$$d = \sum_{i=2}^{n} \frac{(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \quad \text{Durbin-Watson Statistic}$$

Figure C-10. (continued)

Figure C-3 shows the program flow begins in the comparability test section of the program by calculating the t-value for the approved and proposed method means, the t-test critical region and the associated table t-value. These values are written out in section 9a of the output (Figure C-4). If the calculated t-value (t_m) is less than or equal to the table t-value (t_{table}) the methods are comparable and the appropriate caption is written out (Figure C-4 (9a)). If the two methods are found to be non-comparable, the means are tested to see which is larger, the approved mean or the proposed mean. The appropriate caption is written out. The program flow at program locations 108, 109 and 110, also keeps track of which decisions are made for the INFORM summary (Figure C-3).

Regression analysis is the second test in program flow to evaluate method comparability. The formulae for regression are presented in Figure C-10. A general form of the model equation is given along with equations to calculate regression equation slope and intercept, their confidence intervals, the coefficient of determination and the Durbin-Watson Statistic.

The regression analysis is used to evaluate comparability using the theory that data from comparable methods when applied to a regression analysis would yield a slope of one, an intercept of zero, and a coefficient of determination of one. The critical test for the slope is if the 95% confidence interval contains the value one. The critical test for the intercept is if the 95% confidence interval contains the value zero.

The Durbin-Watson value (d-value) is used to detect correlation in the residual terms. Highly correlated residuals indicate the regression model was inappropriate for data evaluation. A d-value of two indicates that there was no correlation for residuals. Figure C-11 below depicts the three critical regions of the Durbin-Watson statistic.

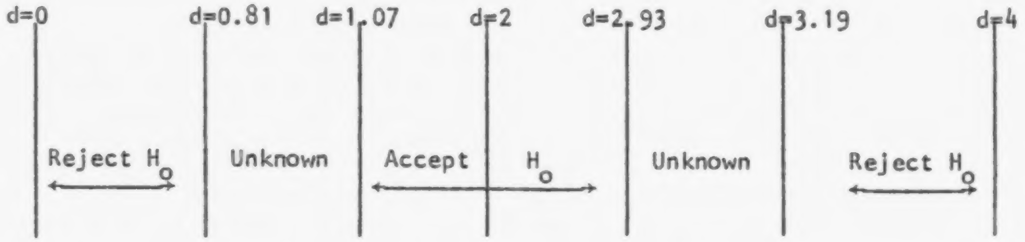


Figure C-11. Critical regions of the Durbin-Watson statistic.

Table C-2 supplies the d-critical values. Only the values on the left hand side of Figure C-11 can be obtained directly from Table C-2. The values on the right hand side can be obtained by subtracting these on the left from four, as follows:

1. $\begin{array}{r} 4.00 \\ \underline{-0.81} \\ 3.19 \end{array}$

$$\begin{array}{lr} 2. & 4.00 \\ & -1.07 \\ \hline & 2.93 \end{array}$$

For purposes of programming, Table C-2 will have to be programmed as a data array from which the two lower level numbers can be obtained for calculating the two higher level numbers.

However, for $n < 15$ the following equations will have to be used to determine the two lower level numbers d_L and d_u:

$$d_L = 0.49 + 0.0223\ n$$

$$d_u = 0.87 + 0.0137\ n$$

It will be up to the staff statistician to determine if the d-values which fall in the unknown region are indicative of serial correlation to the extent that it affects the regression model. However, for the remainder of the discussion on regression analysis programming and the INFORM summary, the regression analysis will be considered appropriate if the d-value falls within either the unknown or acceptable critical regions.

Program flow for the regression analysis section begins by calculating the coefficient of determination, the regression intercept and slope, and the confidence interval for the slope and intercept. The next section of the program calculates the Durbin-Watson Statistic. For this computation, a sorting routine is needed to arrange the approved and proposed method sample means in ascending order. The data is then in the proper sequence for the Durbin-Watson statistic equations in Figure C-10 (7). The sorted $\overline{X}$'s are entered in order into the regression equation (predicted $\overline{Y}_i = B_o + B_i\overline{X}_i$) to obtain the predicted $\overline{Y}$'s. The sorted $\overline{Y}$'s are then subtracted from the corresponding predicted $\overline{Y}$ to obtain e_i (residual or error term). The calculation of predicted $\overline{Y}$ and the subtraction to obtain e_i could be done within the same DO loop. The resulting error terms (e_i) are now in the proper sequence for entry into the Durbin-Watson statistic equation (Figure C-10 (7)). The critical region of the Durbin-Watson statistic is then calculated. Both the d-value and the d-value acceptance range are printed out (Figure C-4(10)). The next portion of the program is a series of decisions about the Durbin-Watson statistic and regression analysis. The first decision block tests to see if the d-value is within the acceptance range. If it is not, the regression analysis was inappropriate and a caption to this effect is written out. A counter for the INFORM System indicating that the regression analysis is inappropriate is then incremented. If the d-value was within the acceptable range, the regression analysis was appropriate and the program continues to the next decision blocks. The program now tests to see if the 95% confidence interval for the slope contains the value one. If not, the approved and proposed methods are non-comparable and a caption to this effect is written out (Figure C-3). The counter for the INFORM System indicating that the regression analysis is inappropriate is then incremented. If the d-value was within the acceptable range, the regression analysis

TABLE C-2. THE DISTRIBUTION OF DURBIN-WATSON d
1 PERCENT SIGNIFICANCE POINTS OF d_L AND d_U (C-1)

n	k=1	
	d_L	d_U
15	0.81	1.07
16	0.84	1.09
17	0.87	1.10
18	0.90	1.12
19	0.93	1.13
20	0.95	1.15
21	0.97	1.16
22	1.00	1.17
23	1.02	1.19
24	1.04	1.20
25	1.05	1.21
26	1.07	1.22
27	1.09	1.23
28	1.10	1.24
29	1.12	1.25
30	1.13	1.26
31	1.15	1.27
32	1.16	1.28
33	1.17	1.29
34	1.18	1.30
35	1.19	1.31
36	1.21	1.32
37	1.22	1.32
38	1.23	1.33
39	1.24	1.34
40	1.25	1.34
45	1.29	1.38
50	1.32	1.40
55	1.36	1.43
60	1.38	1.45
65	1.41	1.47
70	1.43	1.49
75	1.45	1.50
80	1.47	1.52
85	1.48	1.53
90	1.50	1.54
95	1.51	1.55
100	1.52	1.56

was appropriate and the program continues to the next decision blocks. The program now tests to see if the 95% confidence interval for the slope contains the value one. If not, the approved and proposed methods are non-comparable and a caption to this effect is written out (Figure C-3). The counter for the INFORM System indicating that the regression analysis is inappropriate is then incremented. If the d-value was within the acceptable range, the regression analysis was appropriate and the program continues to the next decision blocks. The program now tests to see if the 95% confidence interval for the slope contains the value one. If not, the approved and proposed methods are non-comparable and a caption to this effect is written out (Figure C-3). The counter for the INFORM System statistical summary for non-comparable is also incremented.

If the slope does pass the test, the next decision block is encountered. The program tests to see if the 95% confidence interval for the intercept contains the value zero. If not, the approved and proposed methods are non-comparable and a caption to this effect is written out (Figure C-3). The counter for the INFORM System statistical summary for non-comparable is also incremented. If both the slope and intercept pass the test, the approved and proposed methods are comparable, a caption to this effect is written out, and the appropriate counter for the INFORM statistical summary is incremented (Figure C-3).

Once the comparability tests are completed, a scattergram is printed out. Printing can be done on a peripheral printer (CALCOMP printer) or as part of the program output (line-printer). A scattergram is a graph of data points based on two variables producing a two dimensional scatter diagram. For the equivalency data the x-axis would be the approved method values and the y-axis would be the proposed method values. A plotting routine that can scale both axes will be required.

An example of a scattergram is presented in Figure C-12. This scattergram was created using a "canned" statistical program called "SAS-76", a proprietary program developed by the SAS Institute Inc. The x-axis is the approved method values (AP) and the y-axis is the proposed method values (PR). The regression equation is plotted using stars (*). This scattergram provides a visual aid for analyzing the scatter of actual data points around a line defined by the regression equation developed to detect comparability between the approved and proposed methods.

Once all the sources have been processed, the information going to the INFORM system statistical summary is computed and/or written out to a temporary file where it can be accessed by the INFORM system.

Output

Two types of output will be generated by statistical program;

1. Line printer - numerical results decisions from the statistical tests and a scattergram.

2. INFORM system - summary of the statistical test results.

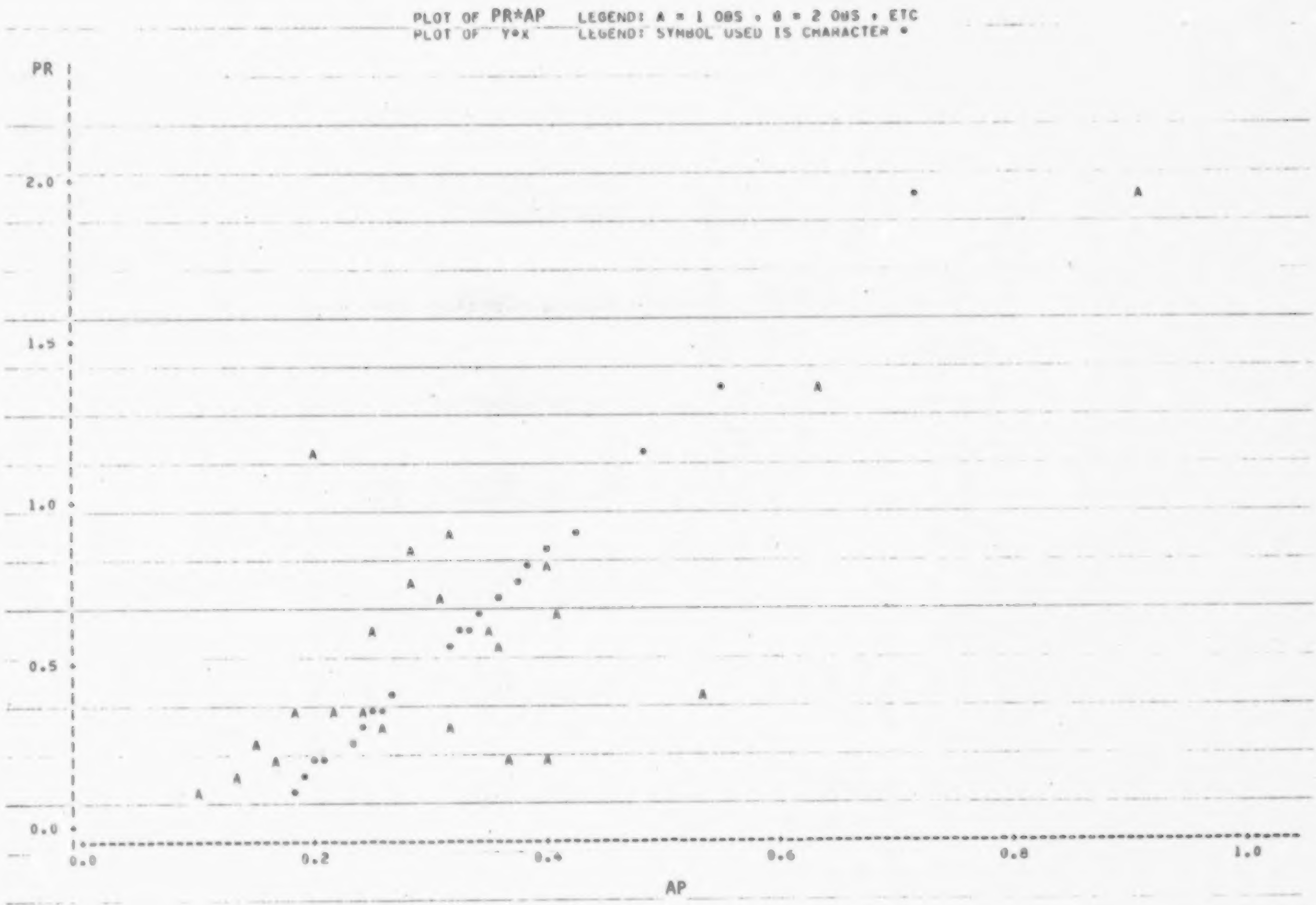


Figure C-12. Example of equivalency data scattergram.

Line Printer--
A sample of output from the statistical program to the line printer is summarized in Figure C-4. The sample output presented is for one source only. For multiple sources, sections 3 through 10 including scattergrams would be duplicated for each source entered into the statistical program. The example of output in Figure C-4 does not present all possible output data and captions. A listing of all possibilities is as follows:

1. Application name and application number - appear first to facilitate identification and expedite filing procedures.
2. Title
    a. Day, date and year.
    b. Parameter being tested.
    c. Type of variance - nationwide or limited.
    d. Type of regulation - NIPDWR of NPDES.
3. Computer listing of input data.
    a. Source number.
    b. Sample number.
    c. Approved method - observations #1 and #2.
    d. Proposed method - observations #1 and #2.
    e. Diluted, blended or spiked sample.
    f. Sample numbers used to generate additional samples.
    g. Composition of generated samples.
    h. Test units - mg/1, ppb, %, etc.
4. Approved and proposed sample means.
5. Approved and proposed source means.
6. Precision
    a. Table t-value.
    b. Calculated t-value (t_R)
    c. Critical t-test range.
    d. Caption indicating test results - one of the following should appear per source:

        -"The ranges of the approved and proposed methods are comparable".

        -"The ranges of the approved and proposed methods are not comparable - the proposed method is more precise than the approved method".

        -"The ranges of the approved and proposed methods are not comparable - the proposed method is less precise than the approved method".

7. Accuracy test.
    a. Source number.
    b. Sample number.
    c. Amount recovered - approved and proposed method.
    d. Background concentrations - approved and proposed method.
    e. Percent recovery - approved and proposed method.
    f. Sample type - diluted or spiked.
    g. Sample composition - percent diluted or composition of spike.

8. Normality testing.
    a. Skewness - approved and proposed method.
    b. Coefficient of kurtosis - approved and proposed method.
    c. Critical value for skewness - approved and proposed method.
    d. Critical value for kurtosis - approved and proposed method.
    e. Caption indicating test results - one of the following should appear per source:

        -"This set of data is normally distributed - no transformation needed".

        -"This set of data is now normally distributed - log data transformation needed".

        -"This set of data is now normally distributed - reciprocal data transformation needed".

        -"This set of data is now normally distributed - square root data transformation needed".

        -"All transformations used - data are not normally distributed - recheck data".

        -"All transformations used - data are not normally distributed - the statistics which follow have been performed on data which are not normally distributed".

    f. Listing of transformed data.

9. Comparability analysis.
    9a. Paired "t" of sample means.
        a. Table t-value.
        b. Calculated t-value (tm).
        c. Critical test range.
        d. Caption indicating test results - one of the following should appear per source:

            -"The means of the approved and proposed methods are comparable".

-"The means of the approved and proposed methods are not comparable - the approved mean is greater than the proposed mean".

-"The means of the approved and proposed methods are not comparable - the proposed mean is greater than the approved mean".

9b. Regression analysis.
  a. "R" squared value.
  b. Regression intercept value.
  c. Confidence value for intercept.
  d. Range of the confidence interval for intercept.
  e. Regression slope value.
  f. Confidence value for slope.
  g. Range of the confidence interval for slope.

10. Durbin-Watson analysis.
  a. Sorted sample means - approved and proposed method.
  b. Durbin-Watson statistic - d-value.
  c. The acceptable range for Durbin-Watson values.
  d. Caption indicating test results - one of the following should appear per source:

  -"Regression analysis was not appropriate".

  -"Regression analysis shows the approved and proposed methods to be comparable".

  -"Regression analysis shows the approved and proposed methods not to be comparable".

11. Scattergram

INFORM System--
Once the program flow jumps outside the outer program loop, the values for the INFORM system statistical summary are written out to a temporary file. From the temporary file the INFORM system can process the statistical summary data to the appropriate permanent file of the applicant. The statistical summary (Figure C-2) contains the following information:

1. Number of sources.

2. Precision.
  a. Comparable.
  b. Non-comparable.
    -Proposed less precise than approved method.
    -Proposed more precise than approved method.

3. Accuracy.
   a. Approved mean and range.
   b. Proposed mean and range.

4. Comparability.
   a. t-test.
      - Comparable.
      - Non-comparable.
        - Proposed mean greater than approved.
        - Approved mean greater than proposed.

   b. Regression.
      - Comparable.
      - Non-comparable.
      - Regression analysis not applicable.

TEST PLAN

The data presented in section 3 of Figure 4 could be used to check and debug the program after it is developed and coded. The values which appear in section 4 through 10 are actual values which have been hand calculated from the sample data. Several points should be kept in mind while developing and debugging the program:

1. Because the statistics are being run simultaneously on values for the approved and proposed method, it would be advantageous to designate the variables using some form of the letters "APPR" and "PROP" to eliminate any confusion while developing the program.

2. Care must be taken to use absolute values whenever they are indicated either on the flow chart or in the formulae.

3. If all the transformations are tried on the data and none produce values which test normal, an error message to recheck the data will be printed out by the program (Figure C-3). This situation could be due to an error in raw data input such as mispunched or transposed numbers.

REFERENCES

C-1. Huang, David, S., Regression and Economic Methods, John Wiley and Sons, Inc., New York, New York, 1970, pg. 256.

APPENDIX D

APPLICATIONS FOR ALTERNATE PROCEDURE USE
APPLICANTS INSTRUCTION AND FORM

PRELIMINARY APPLICATION FOR ALTERNATE PROCEDURE USE

An "Application for the Use of Alternate Analytical Methods" is enclosed along with instructions for completing the application. When completed, this application and a detailed description of the proposed alternate methods should be sent to the appropriate agency.

National Pollutant Discharge Elimination System (NPDES) Program

If you are requesting a nationwide NPDES variance, your application should be sent in triplicate to:

Director, Environmental Monitoring and Support Laboratory
U.S. Environmental Protection Agency
26 W. St. Clair Street
Cincinnati, Ohio 45268

for review and consideration for approval. Requests for a limited use variance under the NPDES program should be sent to the EPA Regional Administrator (RA) for your region. These may also be forwarded to the Environmental Monitoring and Support Laboratory (EMSL) for review and recommendations, although the RA has the authority to approve limited use variances.

You may expect to be informed within 90 days after the date of submittal of your application as to whether your application was 1) accepted, 2) rejected or 3) that comparative test data is required. For limited use variance results, you can contact your Regional Quality Assurance (QA) Coordinator at any time for information regarding the status of your application. Requests for information on nationwide variance approvals should be sent to the Director, EMSL-Cincinnati.

The amendments to the "Guidelines Establishing Test Procedures for the Analysis of Pollutants" published in the Federal Register on December 1, 1976 (40 CFR 136) provide for the use of procedures other than those listed as "approved" procedures if the applicant can show that the "alternate" procedure produces results which are comparable. Careful completion of this application may eliminate the need to perform the analytical testing required to show whether the proposed alternate procedure produces results comparable to an approved procedure. It is possible that sufficient data may be on file from previous applications regarding this procedure which will allow acceptance or rejection of your application based on a method description alone. If there is insufficient information on file, however, it will be necessary to perform a series of analytical tests on wastewater samples in order to test the method comparability. If such testing is needed, you will receive the appropriate report forms, instructions describing the sample types that should be tested and the manner in which the tests should be performed. Do not start comparison tests before receiving the report forms and the instructions.

National Interim Primary Drinking Water Regulations (NIPDWR) Program

If you are requesting a nationwide variance for the use of an alternate procedure for analysis of drinking water under the NIPDWR program, this application and a copy of the proposed test procedure should be sent in triplicate to:

Director, Environmental Monitoring and Support Laboratory
U.S. Environmental Protection Agency
26 W. St. Clair Street
Cincinnati, Ohio 45268

Based on the review of this application, the Director of EMSL may request the applicant to furnish data from comparative testing. In this case, a set of data reporting forms and instructions for sampling and analysis will be sent to you. On the completion of his review, the Director of EMSL will forward his technical recommendations to the Deputy Assistant Administrator Office of Drinking Water (DAA/ODW) for approval or rejection of the variance applications.

Applications for a limited use variance under the NIPDWR program should be submitted in quadruplicate to the appropriate State agency, if the State has accepted primacy under the Safe Drinking Water Act, or directly to the EPA Regional Administrator if the State involved is a non-primacy state. Applications to State agencies will be forwarded to the EPA along with evaluations and recommendations. Based on the recommendations of the State, EMSL and the DAA/ODW, the Regional Administrator may approve the request for variance, reject the request or ask for additional information (such as comparative testing). If EMSL determines that a proposed alternate procedure has the potential for nationwide use, the requested variance may be upgraded to nationwide status.

Careful completion of this application may save a significant amount of time in the final processing. If no further information is required of the applicant, final action on this application can be expected in about eight weeks after receipt of the application by the Director of EMSL (nationwide application) or the EPA Regional Administrator (limited use application). If further information is required, final action will be delayed until about eight weeks after receipt of this additional information. Note that these time frames do not include the time required by the State to process limited use applications. For information concerning the status of your application, contact the agency to which the application was sent.

The Director of EMSL may require that the applicant submit analytical data comparing the proposed alternate method to a method of analysis listed as an approved procedure in the Federal Register on July 9, 1976 (40 CFR 141). In this case, the appropriate data forms and instructions will be sent to you. Do not start the comparison tests before receiving the report forms.

SUMMARY OF INSTRUCTIONS

NPDES Variance Request

Nationwide request - The alternate procedure is useful for any monitoring program requiring the use of NPDES approved methods.

Step 1 - Complete preliminary application form in triplicate.

Step 2 - Prepare description of proposed analytical method following format utilized in the current edition of EPA publication entitled "Manual of Methods for Chemical Analysis of Water and Wastes". Attach a copy to each application form.

Step 3 - Send to:
Director, Environmental Monitoring and Support Laboratory
U.S. Environmental Protection Agency
26 W. St. Clair Street
Cincinnati, Ohio 45268

Limited use request - The alternate procedure is useful only for those situations specified in the application.

Steps 1 and 2 - Same as nationwide NPDES request.

Step 3 - Send to: EPA Regional Administrator in your region.

NIPDWR Variance Request

Nationwide request - Follow same directions as for NPDES nationwide variance request.

Limited use request - Steps 1 and 2 - Same as nationwide NPDES request.

Step 3 - Send to: Appropriate State Agency - if your state has accepted primacy under the Safe Drinking Water Act or EPA Regional Administrator in your Region.

Instructions for the Completion of the Application for the Use of Alternate Analytical Methods

Lines 1, 2 & 3	Enter your company name and address
Line 4	Enter the name and telephone number of the person to whom inquiries concerning this application may be made.
Line 5	Enter the EPA Region and State in which your company is located.
Line 6	Enter the date of submittal.
Lines 7 and 8	Complete line 7 if your request for variance applies to drinking water analysis under the National Interim Primary Drinking Water Regulations (NIPDWR) program. Complete line 8 if your request applies to wastewater analysis under the National Pollutant Discharge Elimination System (NPDES) program.
Line 7	If you are applying for a variance under the NIPDWR program, mark the type of variances you seek: Limited use - The variance is limited geographically or by water supplier. Nationwide - This type of variance allows use of the alternate procedure for NIPDWR compliance monitoring.
Line 8	If you are applying for a variance under the NPDES program, mark the type of variance you seek: Limited use - This type of variance can be a permit-holder type which allows the use of an alternate procedure only for those outfalls covered by one NPDES permit or it may be more general (laboratory type, regionwide). However, it is more limited than a nationwide variance. Nationwide - The alternate procedure can be used for any monitoring program requiring the use of NPDES approved procedures.
Line 9	Briefly describe the scope of the requested limited variance. A limited use variance may restrict the use of an alternate procedure to an EPA Region, a state or a particular laboratory.
Line 10	Identify the parameter or contaminant being measured. Be sure the name accurately describes the analyzed material. For example, don't list "phosphate" when "ortho-phosphate" or "total phosphate" would be more accurate.

Lines 11, 12 and 13	Briefly summarize the proposed alternate method. If the proposed method is described in a publication, include the reference source.
Lines 14 through 20	These lines will be used by the EPA to keep track of the status of your application.
Attachments:	A detailed description of the proposed method must be included with this application. The description should follow the format used in the current edition of EPA *Manual of Methods for Chemical Analysis of Water and Wastes*. Pertinent references and literature citations should be included.

FORM 1

Preliminary Application for the Use of Alternate Analytical Methods

Line No.

1. Applicant Name ____________________
2. Address: ____________________
3. ____________________
4. Contact: ____________________ Telephone: __________
5. EPA Region: ____________________ State: __________
6. Date of Application: ____________________

Type of Variance Requested: (Make only one mark in either line 7 or 8)

7. NIPDWR Program: Limited Use __________ Nationwide __________
8. NPDES Program: Limited Use __________ Nationwide __________
9. If "Limited" is marked, briefly describe the scope of the requested variance: ____________________
10. Parameter or contaminant measured: ____________________

Briefly describe the proposed alternate analytical method:

11. ____________________
12. ____________________
13. ____________________

(Attach a complete description of the proposed alternate method to this form)

Information below this line to be added by EPA only

14. EPA Region: __________ Date Received: __________ Reviewer: __________
15. Application Approved: _____ Disapproved: _____ Need additional data: _____
16. Application forwarded to EMSL for: File: _____ Recommendations: _____
17. Date Received by EMSL: ____________________
18. EMSL Recommendations: Approve: __________ Disapprove: __________
    Testing Required: __________
19. Reviewer: ____________________ Date: __________
20. EMSL File No.: ____________________ EMSL Application No.: __________

DETAILED DATA FORMAT - NPDES NATIONWIDE DATA REQUEST

We do not have sufficient data from the alternate analytical method you propose to use to determine whether this method gives results comparable to the NPDES approved method. It will be necessary to perform analyses on actual samples using both the proposed and approved methods. For a nationwide variance, the proposed and approved methods must be compared using wastewater samples from several different sources. For your application it is recommended that samples be obtained from the following Standard Industrial Classification (SIC) codes for testing:

1. SIC______________________________
2. SIC______________________________
3. SIC______________________________
4. SIC______________________________
5. SIC______________________________
6. SIC______________________________

These sources represent the majority of outfalls monitored for the parameter measured by your proposed alternate method. The Quality Assurance (QA) Coordinator in your Region or your State environmental protection agency may be able to help identify specific industries in your area that have the type of wastewater discharge that should be tested. The necessary forms and directions are included with this packet. When completed, the forms should be submitted to: The Director, Environmental Monitoring and Support
Laboratory
U.S. Environmental Protection Agency
26 W. St. Clair Street
Cincinnati, Ohio 45268

Collection of Samples

Collect six (6) samples from each of the SIC sources listed above. Each waste source may have more than one outfall and the characteristics of the waste from each outfall may vary from hour-to-hour and day-to-day. The sampling location and times should be chosen so that the six samples may be expected to represent the variation in waste strength that occurs at that waste source.

The type of sample taken should be consistent with the NPDES permit requirements for the industry sampled and the parameter to be measured. Generally, these will be either grab samples or 24 hour composite samples. It may be possible to utilize the existing sampling equipment used to routinely obtain samples for the NPDES permit.

The volume of sample collected should be sufficient to perform the required analyses and to prepare the additional samples described later in these instructions. It is recommended that the sample size be at least 15 times larger than the volume needed for one analysis. Accepted sample preservation techniques and holding times required in the approved method must be followed.

Analyses of Samples

To establish comparability, it is necessary to analyze each sample using the proposed alternate procedure and a procedure approved for NPDES compliance monitoring. If there is more than one approved analytical method for the parameter being analyzed, choose the method that is normally used for the wastewaters sampled. The same approved method should be used throughout the comparability testing.

The analytical work must be performed by at least two different laboratories when applying for a nationwide variance. It is recommended that one laboratory analyze all the samples from one or more SIC source and the other laboratory (ies) analyze the remaining samples. All laboratories should perform both the proposed method and the approved method on the samples they analyze. Sample collection and solicitation of other participating laboratories is the responsibility of the applicant.

In order to properly compare the results of the proposed alternate method to the approved method, it is desirable that the sample concentrations be evenly distributed throughout the working range of the proposed alternate method. It is highly unlikely that the six samples obtained from each waste source will have such a concentration distribution. For this reason, six additional samples must be prepared by spiking, blending, and diluting samples selected from the six natural samples obtained from each SIC source. (Do not mix samples from different SIC sources). It is recommended that the six natural samples from each source be analyzed before preparing the additional samples. This will aid in the selection of samples to be used in the preparation of the additional samples. Two of the additional samples should be prepared by spiking two of the natural samples (spiking is the addition of a known concentration of the material being analyzed). The other four samples can be prepared by blending or diluting selected natural samples. The additional samples should be prepared so that the concentrations of these samples cover the portions of the working range of the proposed alternative method not covered by the six natural samples. An example of a method of preparing the six additional samples is included with this information.

All samples must be analyzed in duplicate using each analytical method. This will result in four analyses per sample and 48 analyses per SIC source. If samples from five SIC sources are analyzed, a total of 240 analyses are required; 120 analyses by each method.

In order to minimize the chance of systematic variations influencing the results of this comparability testing, it is recommended that each laboratory randomize their analyses of these samples. Among the suggestions to accomplish this are the following:

1. If more than one analyst is involved, each analyst should perform

analyses by both methods.

2. Perform both methods of analyses at the same time or randomize the order of analysis to minimize the effect of time on the results.

3. If the same type of equipment is required for both methods, either use the same piece of equipment for both methods or randomize the use of multiple pieces of equipment between the two methods.

Failure to randomize or control these variables may result in the determination that the two methods do not produce comparable results when, in fact, the difference in results may have been due to differences in analysts, sample age (time), or equipment operation.

Reporting of Data

One data reporting form for each of the SIC sources to be tested has been included in this packet. These forms are to be filled out by the laboratory performing the analyses. Please forward the appropriate number of these forms to the laboratories involved. When completed, the forms should be returned to you (the applicant) for submittal to the Director, Environmental Monitoring and Support Laboratory (EMSL) at the address listed on the first page of these instructions. Any questions concerning the status of an application should be directed to the Director, EMSL-Cincinnati.

A copy of the "Guide for Completing Data Report Form" included in this packet should be sent to other laboratories participating in these tests. An example of a completed data report form is included. In this example, the method of reporting the preparation of the additional samples described previously is illustrated.

Statistical Evaluation Procedure

The data which are submitted will be evaluated using a statistical approach developed especially for NPDES method comparisons. The procedure has been computerized and the input will be taken directly from your application forms.

The statistical tests will involve an initial calculation of the mean of the duplicate and then a comparison of the ranges for each of the analytical methods. This will allow a general comparison of the precision of the two methods.

The data will then be tested for normality using a Sample Moments Test. Appropriate transforms will be applied to the data in order to make them normal, if necessary. Two important determinations will be made at this point.

1. If, after transformation, one distribution is normal and another is not, then the data sets will be considered to be non-comparable and the request for variance may be rejected.

2. If the data cannot be transformed to a normal distribution, then

they may also be rejected at this point. This determination will be made by the data analysts, who have a choice of whether to continue in the evaluation or return the data to the applicant with a suggestion to rerun certain analyses.

After the normality evaluation is completed, the average of the duplicates for each method will be evaluated using a paired t test and a regression analyses. Both tests will be compared to results at a 95% confidence interval. These evaluations will be performed for each waste source and for the combined data. The t test will result in a single number for comparison purposes. However, the regression analysis will result in the comparison of the slope, intercept and r value (correlation coefficient) to their critical values. In addition, the autocorrelation value and the Durbin-Watson statistic will be determined to indicate if undesirable trends are present in the data. A plot of residuals will also aid in checking the rigor of the statistics with respect to the data.

A recommendation of acceptance or rejection of the alternate method will be made by the Director of EMSL and transmitted to the appropriate EPA office. You will be notified of the decision within 90 days after submittal of the data.

EXAMPLE OF METHOD OF PREPARATION OF THE REQUIRED ADDITIONAL SAMPLES

Assume that the six natural samples had the following concentrations:

Sample No.	Approved Method Data (mg/l)		Proposed Method Data (mg/l)	
1	138	142	140	142
2	38	44	38	42
3	63	58	59	62
4	120	118	121	117
5	1,500	1,450	1,450	1,440
6	218	210	211	220

Also, assume that the working range of the proposed alternate analytical method is 20 to 2,000 mg/l. Considering the concentration of the six natural samples, it would appear desirable that the six additional samples include one sample with a concentration near the lower limit of the working range, one sample near the upper limit and several samples between 210 and 1,500 mg/l. Based on these observations, the following procedure for the preparation of the six additional samples is reasonable:

Sample No.	Sample Preparation	Expected concentration mg/l
7	Dilute sample 2 with an equal volume of deionized water	about 20
8	Mix 1 part sample 4 with 2 parts sample 5	about 1,000
9	Mix 3 parts sample 3 with 1 part sample 5	about 400
10	Mix 1 part sample 1 with 1 part sample 5	about 820
11	Add 500 mg/l of the parameter measured to sample 5	about 2,000
12	Add 400 mg/l of the parameter measured to sample 6	about 600

continued

(continued)

Using these additional samples we are able to:

1. Compare the two methods near the extremes of the proposed alternate method (samples 7 and 11).
2. Compare the two methods at concentrations not covered by the original natural samples (samples 7 through 12).
3. Obtain information as to the accuracy of the two methods (samples 7, 11 and 12).

Guide for Completing the Data Report Form

Line 1	This line should be completed by the EPA office which sent these forms to you.
Line 2	Enter the name of the applicant requesting the variance and the EPA region in which the applicant is located.
Lines 3, 4 and 5	Name and address of the laboratory which analyzed the samples.
Line 6	Name and telephone number of analyst capable of answering questions concerning the analytical results.
Line 7	This entry should be one of the waste sources recommended by the EPA in the letter accompanying these data forms.
Line 8	List amounts and concentrations of preservatives used. Also indicate if sample was maintained at a low temperature.
Line 9	The name of the material analyzed.
Line 10	The actual working range of the method without dilution of concentration.
Line 11	Enter the page number and the name of the publication where the approved method used in these tests is described. This reference must be one of the references listed in the Federal Register designating the appropriate analytical methods (40 CFR 136, Dec. 1, 1976).
Line 12	List the units in which the analytical data is reported. The effective working range entered in line 10 should be in the same units.
Line 13 thru 18	List the results of the duplicate analyses of the natural samples performed by the approved method in columns 1 and 2 and the results from the proposed method in columns 3 and 4. Columns 5, 6 and 7 should not be filled in for lines 13 through 18. Be sure the number of significant figures used are consistent with the precision of the method.
Line 19 thru 24	In the same manner, list the results of the duplicate analyses of the prepared samples in columns 1, 2, 3 and 4.

The method used to prepare the additional samples should be indicated in column 5 by entering:

S if the sample was prepared by spiking,
D if the sample was prepared by dilution,
B if the sample was prepared by blending.

The sample number(s) used to prepare the additional samples should be entered in column 6. If the sample was prepared by diluting or spiking, only one number would be entered. If the sample was prepared by blending, two or more numbers will be entered (separate the sample numbers with a comma). All numbers entered in column 6 should be between 1 and 6.

Enter the following information in column 7:

> diluted samples - Enter the fraction of the diluted sample that is dilution water.
>
> blended samples - Enter the fractions corresponding to the relative amounts of the natural samples used to form the new sample. List the fractions in the same order as the sample numbers listed in column 6.
>
> spiked samples - Indicate the expected change in concentration resulting from the addition of the known amount of chemical. The units should be the same as listed in line 12.

Line 25	Identify the material used to spike the sample. Give a chemical formula, if possible.
Lines 26, 27 and 28	Add any comments about the analysis that may aid in the interpretation of the data. In particular, note any modifications made to either procedure for the tests listed on this form. Any sample pretreatment (such as homogenation or digestion) should be described. Continue comments on an additional sheet if necessary.

EXAMPLE OF COMPLETED DATA REPORT

Comparative Test Data for Evaluating Method Equivalency

1. Application No.: 12345 File No.: 17-86B
2. Applicant Name: ABC Company Region: 5
3. Name of Laboratory Performing Analyses: XYZ Laboratory
4. Address: 789 Main Street
5. Walleye, WI 53000
6. Analyst in Charge: Joe Smith Telephone No.: 123/456-7890
7. Waste Source Tested: Outfall 001
8. Preservatives Used: H_2SO_4 to pH less than 2.
9. Analytical Parameter Tested: COD
10. Effective Working Range of Proposed Method: 20 to 2000 mg/l
11. Approved Method Reference: 14th Ed. Standard Methods, p. 550
12. Units: mg/l

Analytical Results

	Sample Number	Approved Methods Data		Proposed Method Data		Preparation of additional samples		
						Type of Prep	Samples Used	Composition of New Samples
		(col 1)	(col 2)	(col 3)	(col 4)	(col 5)	(col 6)	(col 7)
13.	1	138	142	140	142			
14.	2	38	44	38	42			
15.	3	63	58	59	62			
16.	4	120	118	121	117			
17.	5	1,500	1,450	1,450	1,440			
18.	6	218	210	211	220			
19.	7	20	24	17	19	D	2	0.5
20.	8	980	990	990	1,010	B	4,5	0.33, 0.67
21.	9	370	380	380	360	B	3,5	0.75, 0.25
22.	10	820	810	880	800	B	1,5	0.50, 0.50
23.	11	2,000	1,950	1,950	1,940	S	5	500
24.	12	618	610	610	600	S	6	400

25. Material used to spike samples: Potassium acid phthalate - $KHC_8H_4O_4$.
26. Comments: Sample chloride concentrations ranged from 10 to 50 mg/l
27. Sample 5 was homogenized before analysis.

DETAILED DATA FORMAT - NPDES LIMITED USE VARIANCE REQUEST

We do not have enough data from the alternate analytical method you propose to use to determine whether this method gives results comparable to the NPDES approved method. It will be necessary to perform analyses on actual samples using both the proposed and approved methods. For your limited variance application, the proposed and approved methods should be compared using wastewater samples obtained from the following waste sources:

1.____

2.____

3.____

4.____

5 ____

6.____

These sources have been chosed specifically to reflect the type of variance which you requested. If you have problems, the Quality Assurance (QA) Coordinator in your Region or your State environmental protection agency may be able to help identify specific discharges in your area that have the type of wastewater that should be tested. The necessary forms and directions are included with this user's packet. When completed, these forms should be sent to:

____

____

____

Collection of Samples

Collect six (6) samples from each of the waste sources listed above. If a general waste source is specified, it may have more than one outfall and the characteristics of the waste from each outfall may vary from hour-to-hour and day-to-day. The sampling location and times should be chosen so that the six samples may be expected to represent the variation in waste strength that occurs at that waste source.

When a single outfall is tested, grab samples may be necessary to achieve the desirable variations in parameter concentration. Otherwise the type of sample taken should be consistent with the NPDES permit requirements for the industry sampled and the parameter to be measured. Generally, these will be either grab samples or 24 hour composite samples. It may be possible to utilize the existing equipment already available to obtain samples for the NPDES permit program.

The volume of sample collected should be sufficient to perform the required analyses and to prepare the additional samples described later in these instructions. It is recommended that the sample size be at least 15 times larger than the volume needed for one analysis. Accepted sample preservation techniques and holding times required in the approved methods must be followed. Sample collection is the responsibility of the applicant.

Analyses of Samples

To establish comparability, it is necessary to analyze each sample using the proposed alternate procedure and a procedure approved for NPDES compliance monitoring. If there is more than one approved analytical method for the parameter being analyzed, choose the method that is normally used for the wastewaters sampled. The same approved method should be used throughout the comparability testing.

In order to properly compare the results of the proposed alternate method to the approved method, it is desirable that the sample concentrations be evenly distributed throughout the working range of the proposed alternate method. It is highly unlikely that the six samples obtained from each waste source will have such a concentration distribution. For this reason, six additional samples must be prepared by spiking, blending, and diluting samples selected from the six natural samples obtained from each source. (Do not mix samples from different sources). It is recommended that the six natural samples from each source be analyzed before preparing the additional samples. This will aid in the selection of samples to be used in the preparation of the additional samples. Two of the additional samples should be prepared by spiking two of the natural samples (spiking is the addition of a known concentration of the material being analyzed). The other four samples can be prepared by blending or diluting selected natural samples. The additional samples should be prepared so that the concentrations of these samples cover the portions of the working range of the proposed alternative method not covered by the six natural samples. An example of a method of preparing the six additional samples is included with this packet. All samples must be analyzed in duplicate using each analytical method. This will result in four analyses per sample and 48 analyses per waste source. If samples from four waste sources are analyzed, a total of 192 analyses are required; 96 analyses by each method.

In order to minimize the chance of systematic variations influencing the results of this comparability testing, it is recommended that the laboratory randomize the analyses of these samples. Among the suggestions to accomplish this are the following:

1. If more than one analyst is involved, each analyst should perform analyses by both methods.

2. Perform both methods of analyses at the same time or randomize the order of analysis to minimize the effect of time on the results.

3. If the same type of equipment is required for both methods, either use the same piece of equipment for both methods or randomize the

use of multiple pieces of equipment between the two methods.

Failure to randomize or control these variables may result in the determination that the two methods do not produce comparable results when, in fact, the difference in results may have been due to differences in analysts, sample age (time), or equipment operation.

Reporting of Data

One data reporting form for each of the waste sources to be tested has been included in this packet. These forms are to be filled out by the laboratory performing the analyses. When completed, the forms should be returned to you (the applicant) for submittal to the agency listed on the first page of these instructions. All the data should be sent at the same time. Any questions concerning the status of an application should be directed to the same agency.

An example of a completed data report form is included. In this example, the method of reporting the preparation of the additional samples described previously is illustrated.

Statistical Evaluation Procedure

The submitted data will be evaluated using a statistical approach developed especially for NPDES method comparisons. The procedure has been computerized and the input will be taken directly from your application forms.

The statistical tests will involve an initial calculation of the mean of the duplicate and then a comparison of the ranges for each of the analytical methods. This will allow a general comparison of the precision of the two methods.

The data will then be tested for normality using a Sample Moments Test. Appropriate transforms will be applied to the data in order to make the data normal, if necessary. Two important determinations will be made at this point.

1. If, after transformation, one distribution is normal and another is not, then the data sets will be considered to be non-comparable and the request for variance may be rejected.
2. If the data cannot be transformed to a normal distribution, then they may also be rejected at this point. This determination will be made by the data analysts, who have a choice of whether to continue in the evaluation or return the data to the applicant with a suggestion to rerun certain analyses.

After the normality evaluation is completed, the average of the duplicates for each method will be evaluated using a paired t test and a regression analyses. Both tests will be compared to results at a 95% confidence interval. These evaluations will be performed for each waste source and for the combined data. The t test will result in a single number for comparison purposes. However, the regression analysis will result in the comparison of the slope, intercept and r value (correlation coefficient) to their critical values. In addition, the autocorrelation value and the Durbin-Watson statistic will be determined to indicate if undesirable trends are present in the data. A plot of residuals will also aid in checking the rigor of the statistics with respect to the data.

The final determination of acceptance or rejection of the alternate method will be made by the Regional Administrator as a result of the technical recommendations. You will be notified of the decision within 90 days after submittal of the data.

EXAMPLE OF METHOD OF PREPARATION OF THE REQUIRED ADDITIONAL SAMPLES

Assume that the six natural samples had the following concentrations:

Sample No.	Approved Method Data (mg/l)		Proposed Method Data (mg/l)	
1	138	142	140	142
2	38	44	38	42
3	63	58	59	62
4	120	118	121	117
5	1,500	1,450	1,450	1,440
6	218	210	211	220

Also, assume that the working range of the proposed alternate analytical method is 20 to 2,000 mg/l. Considering the concentration of the six natural samples, it would appear desirable that the six additional samples include one sample with a concentration near the lower limit of the working range, one sample near the upper limit and several samples between 210 and 1,500 mg/l. Based on these observations, the following procedure for the preparation of the six additional samples is reasonable:

Sample No.	Sample Preparation	Expected concentration mg/l
7	Dilute sample 2 with an equal volume of deionized water	about 20
8	Mix 1 part sample 4 with 2 parts sample 5	about 1,000
9	Mix 3 parts sample 3 with 1 part sample 5	about 400
10	Mix 1 part sample 1 with 1 part sample 5	about 820
11	Add 500 mg/l of the parameter measured to sample 5	about 2,000
12	Add 400 mg/l of the parameter measured to sample 6	about 600

(continued)

(continued)

Using these additional samples we are able to:

1. Compare the two methods near the extremes of the proposed alternate method (samples 7 and 11).
2. Compare the two methods at concentrations not covered by the original natural samples (samples 7 through 12).
3. Obtain information as to the accuracy of the two methods (samples 7, 11 and 12).

Guide for Completing the Data Report Form

Line 1	This line should be completed by the EPA office which sent these forms to you.
Line 2	Enter the name of the applicant requesting the variance and the EPA region in which the applicant is located.
Lines 3, 4 and 5	Name and address of the laboratory which analyzed the samples.
Line 6	Name and telephone number of analyst capable of answering questions concerning the analytical results.
Line 7	This entry should be one of the waste sources recommended by the EPA in the letter accompanying these data forms.
Line 8	List amounts and concentrations of preservatives used. Also indicate if sample was maintained at a low temperature.
Line 9	The name of the material analyzed.
Line 10	The actual working range of the method without dilution of concentration.
Line 11	Enter the page number and the name of the publication where the approved method used in these tests is described. This reference must be one of the references listed in the Federal Register designating the appropriate analytical methods (40 CFR 136, Dec. 1, 1976).
Line 12	List the units in which the analytical data is reported. The effective working range entered in line 10 should be in the same units.
Line 13 thru 18	List the results of the duplicate analyses of the natural samples performed by the approved method in columns 1 and 2 and the results from the proposed method in columns 3 and 4. Columns 5, 6 and 7 should not be filled in for lines 13 through 18. Be sure the number of significant figures used are consistent with the precision of the method.
Line 19 thru 24	In the same manner, list the results of the duplicate analyses of the prepared samples in columns 1, 2, 3 and 4.

The method used to prepare the additional samples should be indicated in column 5 by entering:

S if the sample was prepared by spiking,
D if the sample was prepared by dilution,
B if the sample was prepared by blending.

The sample number(s) used to prepare the additional samples should be entered in column 6. If the sample was prepared by diluting or spiking, only one number would be entered. If the sample was prepared by blending, two or more numbers will be entered (separate the sample numbers with a comma). All numbers entered in column 6 should be between 1 and 6.

Enter the following information in column 7:

diluted samples - Enter the fraction of the diluted sample that is dilution water.

blended samples - Enter the fractions corresponding to the relative amounts of the natural samples used to form the new sample. List the fractions in the same order as the sample numbers listed in column 6.

spiked samples - Indicate the expected change in concentration resulting from the addition of the known amount of chemical. The units should be the same as listed in line 12.

Line 25	Identify the material used to spike the sample. Give a chemical formula, if possible.
Lines 26, 27 and 28	Add any comments about the analysis that may aid in the interpretation of the data. In particular, note any modifications made to either procedure for the tests listed on this form. Any sample pretreatment (such as homogenation or digestion) should be described. Continue comments on an additional sheet if necessary.

EXAMPLE OF COMPLETED DATA REPORT

Comparative Test Data for Evaluating Method Equivalency

1. Application No.: 12345 File No.: 17-86B
2. Applicant Name: ABC Company Region: 5
3. Name of Laboratory Performing Analyses: XYZ Laboratory
4. Address: 789 Main Street
5. Walleye, WI 53000
6. Analyst in Charge: Joe Smith Telephone No.: 123/456-7890
7. Waste Source Tested: SIC 2911 - Petroleum and Refining
8. Preservatives Used: H_2SO_4 to pH less than 2.
9. Analytical Parameter Tested: COD
10. Effective Working Range of Proposed Method: 20 to 2000 mg/l
11. Approved Method Reference: 14th Ed. Standard Methods, p. 550
12. Units: mg/l

Analytical Results

	Sample Number	Approved Methods Data		Proposed Method Data		Preparation of additional samples		
						Type of Prep	Samples Used	Composition of New Samples
		(col 1)	(col 2)	(col 3)	(col 4)	(col 5)	(col 6)	(col 7)
13.	1	138	142	140	142			
14.	2	38	44	38	42			
15.	3	63	58	59	62			
16.	4	120	118	121	117			
17.	5	1,500	1,450	1,450	1,440			
18.	6	218	210	211	220			
19.	7	20	24	1.	19	D	2	0.5
20.	8	980	990	990	1,010	B	4,5	0.33, 0.67
21.	9	370	380	380	360	B	3,5	0.75, 0.25
22.	10	820	810	880	800	B	1,5	0.50, 0.50
23.	11	2,000	1,950	1,950	1,940	S	5	500
24.	12	618	610	610	600	S	6	400

25. Material used to spike samples: Potassium acid phthalate - $KHC_8H_4O_4$.
26. Comments: Sample chloride concentrations ranged from 10 to 50 mg/l.
27. Sample 5 was homogenized before analysis.

DETAILED DATA FORMAT - NIPDWR NATIONWIDE VARIANCE REQUEST

We do not have sufficient data from the alternate analytical method you propose to use to determine whether this method gives results comparable to the NIPDWR approved method. It will be necessary to perform analyses on actual samples using both the proposed and approved methods. For a nationwide variance, the proposed and approved methods should be compared using water samples from several different sources. For your application, it is recommended that samples be obtained from the following sources for testing:

1. ______________________________

2. ______________________________

3. ______________________________

4. ______________________________

5. ______________________________

6. ______________________________

The necessary forms and directions are included with this packet. When completed, the forms should be submitted to:

The Director, Environmental Monitoring and Support Laboratory
U.S. Environmental Protection Agency
26 W. St. Clair Street
Cincinnati, Ohio 45268

Collection of Samples

Collect six (6) samples from each of the sources listed above. The sampling location and times should be chosen so that the six samples may be expected to represent the variation in strength that occurs at that source.

The type of sample taken should be consistent with the NIPDWR program requirements for the contaminant to be measured.

The volume of sample collected should be sufficient to perform the required analyses and to prepare the additional samples described later in these instructions. It is recommended that the sample size be at least 15 times larger than the volume needed for one analysis. Sample preservation techniques and holding times prescribed for use in the approved NIPDWR method must be used.

Analyses of Samples

To establish comparability, it is necessary to analyze each sample using the proposed alternate procedure and the procedure approved for NIPDWR compliance monitoring. If there is more than one approved analytical method the difference in results may have been due to difference in analysts, sample age (time), or equipment operation.

Reporting of Data

One data reporting form for each of the water sources to be tested has been included in this packet. When completed, the forms should be returned to you (the applicant) for submittal to the Director, Environmental Monitoring and Support Laboratory, (EMSL) at the address listed on the first page of these instructions. All the data should be sent at the same time. Any questions concerning the status of an application should be directed to the same address.

An example of a completed data report form is included. In this example, the method of reporting the preparation of the additional samples described previously is illustrated.

Statistical Evaluation Procedure

The data which is submitted will be evaluated using a statistical approach developed especially for the Environmental Monitoring and Support Laboratory (EMSL) Cincinnati for analytical method comparisons. The procedure has been computerized and the input will be taken directly from your application forms.

The statistical tests will involve an initial calculation of the mean of the duplicates and then a comparison of the ranges for each of the analytical methods. This will allow a general comparison of the precision of the two methods.

The data will then be tested for normality using a Sample Moments Test. Appropriate transforms will be applied to the data in order to make it normal, if necessary. Two important determinations will be made at this point.

1. If, after transformation, one distribution is normal and another is not, then the data sets will be considered to be non-comparable and the request for alternate method may be rejected.
2. If the data cannot be transformed to a normal distribution, then if may also be rejected at this point. This determination will be made by the data analyst, who has a choice of whether to continue in the evaluation or return the data to the applicant with a suggestion to rerun certain analyses.

After the normality evaluation is completed, the average of the duplicates for each method will be evaluated using a paired t test and a regression analyses. Both tests will be compared to results at a 95% confidence interval. These evaluations will be performed for each waste source and for the combined data. The t test will result in a single number for comparison purposes. However, the regression analysis will result in the comparison of the slope, intercept and r value (correlation coefficient) to their critical values. In addition, the autocorrelation value and the Durbin-Watson statistic will be determined to indicate if undesirable trends are present in the data. A plot of residuals will also aid in checking the rigor of the statistics with respect to the data.

The final determination of acceptance or rejection of the alternate method will be made by the Deputy Assistant Administrator Office of Drinking Water after consideration of the EMSL technical recommendations and other factors.

EXAMPLE OF METHOD OF PREPARATION OF THE REQUIRED ADDITIONAL SAMPLES

Assume that the six natural samples had the following concentrations:

Sample No.	Approved method data (mg/l)		Proposed method data (mg/l)	
1	0.012	0.010	0.010	0.008
2	0.016	0.020	0.022	0.018
3	0.014	0.016	0.014	0.014
4	0.018	0.016	0.016	0.018
5	0.010	0.012	0.010	0.010
6	0.026	0.024	0.028	0.024

Also, assume that the maximum contaminant level is 0.05 mg/l and the detection limit is 0.001 mg/l. Because the original six samples have concentrations substantially less than the maximum contaminant level, preparation of four of the additional samples by spiking appears to be appropriate. For this example, samples are also prepared by dilution and blending to illustrate these methods. The following procedure might be used for the preparation of the six additional samples:

Sample No.	Sample Preservation	Expected concentration mg/l
7	Add 0.030 mg/l of the contaminant measured to sample 6	0.055
8	Add 0.030 mg/l of the contaminant measured to sample 2	0.050
9	Add 0.020 mg/l of the contaminant measured to sample 3	0.040
10	Add 0.020 mg/l of the contaminant measured to sample 5	0.030
11	Blend 1 part sample 1 with 2 sample 6	0.020
12	Dilute sample 1 with an equal volume of reagent grade water	0.006

Guide for Completing the Data Report Form

Line 1	This line should be completed by the EPA office which sent these forms to you.
Line 2	Enter the name of the applicant requesting the variance and the EPA region in which the applicant is located.
Lines 3, 4 and 5	Name and address of the laboratory which analyzed the samples.
Line 6	Name and telephone number of analyst capable of answering questions concerning the analytical results.
Line 7	This entry should be one of the water sources recommended by the EPA in the letter accompanying these data forms.
Line 8	List amounts and concentrations of preservatives used. Also indicate if sample was maintained at a low temperature.
Line 9	The name of the material analyzed.
Line 10	The actual working range of the method without dilution or concentration.
Line 11	Enter the page number and the name of the publication where the approved method used in these tests is described. This reference must be one of the references listed in the Federal Register designating the appropriate analytical methods (40 CFR 141, July 9, 1976).
Line 12	List the units in which the analytical data is reported. The effective working range entered in line 10 should be in the same units.
Lines 13 thru 18	List the results of the duplicate analyses of the natural samples performed by the approved method in columns 1 and 2 and the results from the proposed method in columns 3 and 4. Columns 5, 6 and 7 should not be filled in for lines 13 through 18. Be sure the number of significant figures used are consistent with the precision of the method.
Lines 19 thru 24	In the same manner, list the results of the duplicate analyses of the prepared samples in columns 1, 2, 3 and 4. The method used to prepare the additional samples should be indicated in column 5 by entering: S if the sample was prepared by spiking, D if the sample was prepared by dilution, B if the sample was prepared by blending.

The sample number(s) used to prepare the additional samples should be entered in column 6. If the sample was prepared by diluting or spiking, only one number would be entered. If the sample was prepared by blending, two or more numbers will be entered (separate the sample numbers with a comma). All numbers entered in column 6 should be between 1 and 6.

Enter the following information in column 7:

diluted samples - Enter the fraction of the diluted sample that is dilution water.

blended samples - Enter the fractions corresponding to the relative amounts of the natural samples used to form the new sample. List the fractions in the same order as the sample numbers listed in column 6.

spiked samples - Indicate the expected change in concentration resulting from the addition of the known amount of chemical. The units should be the same as listed in line 12.

Line 25 Identify the material used to spike the sample. Give a chemical formula, if possible.

Lines 26 27 and 28 Add any comments about the analysis that may aid in the interpretation of the data. In particular, note any modifications made to either procedure for the tests listed on this form. Continue comments on an additional sheet if necessary.

EXAMPLE OF COMPLETED FORM

Comparative Test Data for Evaluating Method Equivalency

1. Application No.: 12345 File No.: 67890
2. Applicant Name: ABC Company Region: 5
3. Name of Laboratory Performing Analyses: XYZ Laboratory
4. Address: 123 Main Street
5. Northern, WI 53000
6. Analyst in Charge: John Smith Telephone No.: 123/456-7890
7. Water Source Tested: Well #1
8. Preservatives Used: 1 ml HNO_3/100 ml
9. Analytical Parameter Tested: Silver
10. Effective Working Range of Proposed Method: 0.002 to 0.1 mg/l
11. Approved Method Reference: 1974 EPA Manual, p. 146; APDC-MIBK extraction
12. Units: mg/l

Analytical Results

	Sample Number	Approved Methods Data		Proposed Method Data		Preparation of additional samples		
						Type of Prep	Samples Used	Composition of New Samples
		(col 1)	(col 2)	(col 3)	(col 4)	(col 5)	(col 6)	(col 7)
13.	1	0.008	0.010	0.010	0.015			
14.	2	0.015	0.020	0.020	0.015			
15.	3	0.015	0.015	0.015	0.015			
16.	4	0.020	0.015	0.015	0.020			
17.	5	0.010	0.010	0.010	0.010			
18.	6	0.032	0.025	0.030	0.025			
19.	7	0.055	0.055	0.060	0.055	5	6	0.030
20.	8	0.045	0.050	0.042	0.045	5	2	0.030
21.	9	0.035	0.040	0.038	0.035	5	3	0.020
22.	10	0.030	0.032	0.030	0.030	5	5	0.020
23.	11	0.022	0.020	0.020	0.020	B	1,6	0.33, 0.67
24.	12	0.005	0.007	0.007	0.005	D	1	0.50

25. Material used to spike samples: $AgNO_3$
26. Comments: Ag concentration of dilution water was <0.001 mg/l.
27.

DATAILED DATA FORMAT-NIPDWR LIMITED USE VARIANCE REQUEST

We do not have sufficient data from the alternate analytical method you propose to use to determine whether this method gives results comparable to the NIPDWR approved method. It will be necessary to perform analyses on actual samples using both the proposed and approved methods. For a limited use variance, the proposed and approved methods should be compared using samples from one or more sources, depending on the scope of the variance. The sources to be sampled are determined on a case-by-case basis. For your application, it is recommended that samples be obtained from the following sources for testing:

1.____

2.____

3.____

4.____

5.____

6.____

The necessary forms and directions are included with this packet. When completed, these forms should be sent to:

____

____

____

Collection of Samples

Collect six (6) samples from each of the sources listed above. The sampling location and times should be chosen so that the six samples may be expected to represent the variation in strength that occurs at that source. The type of sample taken should be consistent with the NIPDWR program requirements for the contaminant to be measured.

The volume of sample collected should be sufficient to perform the required analyses and to prepare the additional samples described later in these instructions. It is recommended that the sample size be at least 15 times larger than the volume needed for one analysis. Sample preservation techniques and holding times prescribed for use in the approved NIPDWR method must be used. Sample collection is the responsibility of the applicant.

Analyses of Samples

To establish comparability, it is necessary to analyze each sample using the proposed alternate procedure and the procedure approved for NIPDWR compliance monitoring. If there is more than one approved analytical method for the contaminant being analyzed, choose the method that is normally used for the waters sampled. The same approved method should be used throughout the comparability testing.

In order to properly compare the results of the proposed alternate method to the approved method, it is desirable that the sample concentrations be evenly distributed throughout the likely range of concentrations for the contaminant measured. It is highly unlikely that the six samples obtained from each source will have such a concentration distribution. For this reason, six additional samples must be prepared by spiking, blending, and diluting samples selected from the six natural samples obtained from each source. (Do not mix samples from different sources). It is recommended that the six natural samples from each source be analyzed before preparing the additional samples. This will aid in the selection of samples to be used in the preparation of the additional samples. The additional samples should be prepared so that the concentrations of these samples cover the portions of the working range of proposed alternative method not covered by the six natural samples. Because drinking water samples from one source tend to have a narrow range of concentrations, it is expected that spiking and diluting will be the most common method of preparing the additional samples. It is recognized that spiking is inappropriate for some contaminants (such as turbidity). In this case, different concentrations can only be obtained by blending or diluting. Dilution should be made with reagent grade water. In some cases, analysis of the dilution water may be desirable. It is recommended that two of the additional samples be prepared so that the final concentration of the contaminant measured is at or near the maximum contaminant level. An example of a method of preparing additional samples is included with this packet.

All samples must be analyzed in duplicate using each analytical method. This will result in four analyses per sample and 48 analyses per source.

In order to minimize the chance of systematic variations influencing the results of this comparability testing, it is recommended that the laboratory randomize the analyses of the samples. Among the suggestions to accomplish this are the following:

1. Each analyst should perform analyses by both methods.
2. Perform both methods of analyses at the same time or randomize the order of analysis to minimize the effect of time on the results.
3. If the same type of equipment is required for both methods, either use the same piece of equipment for both methods or randomize the use of multiple pieces of equipment between the two methods.

Failure to randomize or control these variables may result in the determination that the two methods do not produce comparable results when, in fact, the difference in results may have been due to differences in analysts, sample age (time), or equipment operation.

Reporting of Data

One data reporting form for each of the water sources to be tested has been included in this packet. When completed, the forms should be returned to you (the applicant) for submittal to the agency listed on the first page of these instructions. All the data should be sent at the same time. Any questions concerning the status of an application should be directed to the same agency.

An example of a completed data report form is included. In this example, the method of reporting the preparation of the additional samples described previously is illustrated.

Statistical Evaluation Procedure

The data which is submitted will be evaluated using a statistical approach developed especially for the Environmental Monitoring and Support Laboratory (EMSL) Cincinnati for analytical method comparisons. The procedure has been computerized and the input will be taken directly from your application forms.

The statistical tests will involve an initial calculation of the mean of the duplicate and then a comparison of the ranges for each of the analytical methods. This will allow a general comparison of the precision of the two methods.

The data will then be tested for normality using a Sample Moments Test. Appropriate transforms will be applied to the data in order to make it normal if necessary. Two important determinations will be made at this point.

1. If, after transformation, one distribution is normal and another is not, then the data sets will be considered to be non-comparable and the request for alternate method may be rejected.

2. If the data cannot be transformed to a normal distribution, then it may also be rejected at this point. This determination will be made by the data analyst, who has a choice of whether to continue in the evaluation or return the data to the applicant with a suggestion to rerun certain analyses.

After the normality evaluation is completed, the average of the duplicates for each method will be evaluated using a paired t test and a regression analyses. Both tests will be compared to results at a 95% confidence interval. These evaluations will be performed for each source and for the combined data. The t test will result in a single number for comparison purposes. However, the regression analysis will result in the comparison of the slope, intercept and r value (correlation coefficient) to their critical values. In addition, the autocorrelation value and the Durbin-Watson statistic will be determined to indicate if undesirable trends are present in the data. A plot of residuals will also aid in checking the rigor of the statistics with respect to the data.

The final determination of acceptance or rejection of the alternate method will be made by the EPA Regional Administrator as a result of advice from the State agency, technical recommendations from the Director of EMSL and administrative action by the Deputy Assistant Administrator Office of Drinking Water.

EXAMPLE OF METHOD OF PREPARATION OF THE REQUIRED ADDITIONAL SAMPLES

Assume that the six natural samples had the following concentrations:

Sample No.	Approved method data (mg/l)		Proposed method data (mg/l)	
1	0.008	0.010	0.010	0.015
2	0.015	0.020	0.020	0.015
3	0.015	0.015	0.015	0.015
4	0.020	0.015	0.010	0.010
5	0.010	0.010	0.010	0.010
6	0.032	0.025	0.030	0.025

Also, assume that the maximum contaminant level is 0.05 mg/l and the detection limit is 0.001 mg/l. Because the original six samples have concentrations substantially less than the maximum contaminant level, preparation of four of the additional samples by spiking appears to be appropriate. For this example, samples are also prepared by dilution and blending to illustrate these methods. The following procedure might be used for the preparation of the six additional samples:

Sample No.	Sample Preservation	Expected Concentration mg/l
7	Add 0.030 mg/l of the contaminant measured to sample 6	0.055
8	Add 0.030 mg/l of the contaminant measured to sample 2	0.050
9	Add 0.020 mg/l of the contaminant measured to sample 3	0.040
10	Add 0.020 mg/l of the contaminant measured to sample 5	0.030
11	Blend 1 part sample 1 with 2 parts sample 6	0.020
12	Dilute sample 1 with an equal volume, of reagent grade water	0.006

Guide for Completing the Data Report Form

Line 1	This line should be completed by the EPA office which sent these forms to you.
Line 2	Enter the name of the applicant requesting the variance and the EPA region in which the applicant is located.
Lines 3, 4 and 5	Name and address of the laboratory which analyzed the samples.
Line 6	Name and telephone number of analyst capable of answering questions concerning the analytical results.
Line 7	This entry should be one of the water sources recommended by the EPA in the letter accompanying these data forms.
Line 8	List amounts and concentrations of preservatives used. Also indicate if sample was maintained at a low temperature.
Line 9	The name of the material analyzed.
Line 10	The actual working range of the method without dilution or concentration.
Line 11	Enter the page number and the name of the publication where the approved method used in these tests is described. This reference must be one of the references listed in the Federal Register designating the appropriate analytical methods (40 CFR 141, July 9, 1976).
Line 12	List the units in which the analytical data is reported. The effective working range entered in line 10 should be in the same units.
Lines 13 thru 18	List the results of the duplicate analyses of the natural samples performed by the approved method in columns 1 and 2 and the results from the proposed method in columns 3 and 4. Columns 5, 6 and 7 should not be filled in for lines 13 through 18. Be sure the number of significant figures used are consistent with the precision of the method.
Lines 19 thru 24	In the same manner, list the results of the duplicate analyses of the prepared samples in columns 1, 2, 3 and 4.

The method used to prepare the additional samples should be indicated in column 5 by entering:

S if the sample was prepared by spiking,
D if the sample was prepared by dilution,
B if the sample was prepared by blending.

The sample number(s) used to prepare the additional samples should be entered in column 6. If the sample was prepared by diluting or spiking, only one number would be entered. If the sample was prepared by blending, two or more numbers will be entered (separate the sample numbers with a comma). All numbers entered in column 6 should be between 1 and 6.

Enter the following information in column 7:

> diluted samples - Enter the fraction of the diluted sample that is dilution water.
>
> blended samples - Enter the fractions corresponding to the relative amounts of the natural samples used to form the new sample. List the fractions in the same order as the sample numbers listed in column 6.
>
> spiked samples - Indicate the expected change in concentration resulting from the addition of the known amount of chemical. The units should be the same as listed in line 12.

Line 25	Identify the material used to spike the sample. Give a chemical formula, if possible.
Lines 26, 27 and 28	Add any comments about the analysis that may aid in the interpretation of the data. In particular, note any modifications made to either procedure for the tests listed on this form. Continue comments on an additional sheet if necessary.

EXAMPLE OF COMPLETED FORM

Comparative Test Data for Evaluating Method Equivalency

1. Application No.: 12345 File No.: 67890
2. Applicant Name: ABC Company Region: 5
3. Name of Laboratory Performing Analyses: XYZ Laboratory
4. Address: 123 Main Street
5. Northern, WI 53000
6. Analyst in Charge: John Smith Telephone No.: 123/456-7890
7. Water Source Tested: Well #1
8. Preservatives Used: 1 ml HNO_3/100 ml
9. Analytical Parameter Tested: Silver
10. Effective Working Range of Proposed Method: 0.002 to 0.1 mg/l
11. Approved Method Reference: 1974 EPA Manual, p. 146; APDC-MIBK extraction
12. Units: mg/l

Analytical Results

	Sample Number	Approved Methods Data		Proposed Method Data		Preparation of additional samples		
						Type of Prep	Samples Used	Composition of New Samples
		(col 1)	(col 2)	(col 3)	(col 4)	(col 5)	(col 6)	(col 7)
13.	1	0.008	0.010	0.010	0.015			
14.	2	0.015	0.020	0.020	0.015			
15.	3	0.015	0.015	0.015	0.015			
16.	4	0.020	0.015	0.015	0.020			
17.	5	0.010	0.010	0.010	0.010			
18.	6	0.032	0.025	0.030	0.025			
19.	7	0.055	0.055	0.060	0.055	5	6	0.030
20.	8	0.045	0.050	0.042	0.045	5	2	0.030
21.	9	0.035	0.040	0.038	0.035	5	3	0.020
22.	10	0.030	0.032	0.030	0.030	5	5	0.020
23.	11	0.022	0.020	0.020	0.020	B	1,6	0.33, 0.67
24.	12	0.005	0.007	0.007	0.005	D	1	0.50

25. Material used to spike samples: $AgNO_3$
26. Comments: Ag concentration of dilution water was <0.001 mg/l.
27.

APPENDIX E

RECOMMENDED GUIDELINES FOR ADMINISTRATION OF ALTERNATE METHOD APPLICATIONS

EPA REGIONAL OFFICE

The following information has been developed as suggestions regarding administration of the alternate method protocol developed under this contract.

Initial Requests

Initial requests for the use of an alternate procedure may be received by the Regional office directly from the applicant or from a state agency. A copy of the "Preliminary Application for Alternate Procedure Use" should be sent to the applicant.

NPDES Preliminary Applications

Completed preliminary forms for limited use NPDES variance applications will be submitted to your office. Applications for nationwide variance should be forwarded directly to the Environmental Monitoring and Support Laboratory - Cincinnati (EMSL-Ci) for technical review.

After receipt of an application for a limited use NPDES variance, a determination must be made regarding whether or not additional data is needed prior to approving (or rejecting) an alternate procedure. The following aspects should be considered:

1. Does the proposed method involve a significant departure from an approved procedure? A review of the detailed analytical method discussion included with the application should allow a determination of this factor. If the method is not significantly different, further testing is not necessary.

2. Has a similar method been tested and approved for a similar application? If so, additional analytical work may not be needed or may be reduced in scope.

3. Has a similar method been tested and rejected for a similar application? In this case, the applicant should be advised. If the applicant wishes to test the proposed method for comparability anyway, the appropriate instructions and Data Reporting Form should be furnished to the applicant.

If the proposed method is different than approved techniques and no comparability test work has been done previously, then the applicant must submit sufficient data which will allow a statistical evaluation of comparability. Please complete lines 14 through 17 of the Preliminary Application.

If comparability data is required, the appropriate Detailed Equivalency Application Packet should be sent to the applicant. The waste sources to be sampled by the applicant should be listed on the front page of the instructions. In addition, one data reporting form (labeled "Comparative Test Data for Evaluating Method Equivalency") should be included for each waste source to be sampled. The waste sources selected for sampling should be based on the type of variance requested. In no case, however, should there be more than five waste sources sampled for a limited use variance. The following suggestions have been prepared as an aid in the selection of appropriate waste sources. For more detailed advice, please contact EMSL-Ci.

Permit Holder Type--
Determine the number of permitted outfalls and the volume and composition of the waste streams. The applicant can be required to sample up to four different outfalls if the parameter of interest will be monitored at these locations. If more than four outfalls are pertinent, it is desirable to designate four which have continuous discharge and/or the highest mass loading of the parameter of interest.

Laboratory Type--
If the description of scope defines the major application of the method within a laboratory, a preliminary definition of waste sources to be sampled is possible. If no information is available, then it is recommended that a municipal effluent and up to three major Standard Industrial Classification (SIC) code types be samples - (a list is available from EMSL-Ci.). Another approach would be to specify local industrial or municipal sample sources which the laboratory would service most frequently for the specified parameter.

Region Wide Variance--
As with the laboratory type of variance, it is possible that an understanding of the method scope would allow determination of the waste sources. Otherwise, a municipal effluent and up to three major SIC sources can be specified. Major industrial classes in the region may also be appropriate.

Other Types of Limited Variance--
Contact EMSL-Ci for guidance, if desired.

Receipt of NPDES Detailed Equivalency Application Packet

After the data is returned, please forward it to EMSL-Ci for technical review. After receipt of their recommendation, a determination as to whether the method is to be approved or rejected can be made. Please inform the applicant and EMSL-Ci regarding the decision.

NIPDWR Preliminary Applications

You may receive preliminary applications for the use of an alternate procedure under the NIPDWR program. If a nationwide variance is involved, forward the application to EMSL-Ci for further handling. When a limited use variance is requested, however, a review of the submitted information is needed to establish if comparability data is required for evaluation.

Among the considerations are:

1. Is the method a significant deviation from the standard procedure?
2. Has a similar method been approved for a similar application?
3. Has a similar method been rejected for a similar application?

Lines 14 through 17 on the Preliminary Application form should be completed to record the status of the application.

If it is determined that comparative test data is required, the appropriate Detailed Data Reporting forms and instructions should be sent to the applicant. The water sources to be sampled by the applicant should be listed on the front page of the instructions. To select appropriate sources for sampling, consideration should be given to the sources of water in the area covered by the requested variance. The selected sources should represent the variety of water sources in the area (e.g., ground water, surface water, etc.) and should include the major sources. However, no more than four sources should be required for sampling for a limited variance. One Detailed Data Reporting form should be sent for each water source to be sampled. Please add the identification numbers from line 14 of the Preliminary Application to line 1 of each Data Reporting form.

After receipt of the completed data reports, the information can be forwarded to EMSL-Ci for technical review. EMSL will forward their technical recommendation to the Deputy District Administrator for the Office of Drinking Water (DAA/ODW), and the Regional Administrator can determine whether or not the method is accepted as equivalent. Please inform the appropriate State agency and/or the applicant and EMSL-Ci of the results of your decision. For detailed information regarding the inter-relationships between the State agencies, the Regional Administrator, EMSL-Cincinnati and the Office of Drinking Water, see the memorandum from Victor J. Kimm, Deputy Administrator for Water Supply dated March 10, 1977 entitled "Revised Procedure for 'Approval of Alternative Analytical Methods--National Interim Primary Drinking Water Regulations' ". This memorandum also described the amount of time allowed for each step of the application evaluation.

ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY - CINCINNATI (EMSL-Ci)

The following guidelines have been prepared to aid in the processing of requests for the use of alternate methods under both the NPDES and NIPDWR programs. The Director of EMSL has responsibility for providing technical review and recommendations for limited use and nationwide variance applications.

Receipt of a Preliminary Application

Preliminary applications for both NPDES and NIPDWR nationwide variance requests may be submitted directly to EMSL-Cincinnati. Screening of this information is necessary to determine if comparative test data is required. The following factors should be considered:

1. Does the proposed method involve a significant departure from an approved procedure? A review of the detailed analytical method discussion included with the application should allow a determination of this factor. If the method is not significantly different, comparative testing is not necessary.
2. Has a similar method been tested and approved for a similar application? If so, additional analytical work may not be needed or may be reduced in scope.
3. Has a similar method been tested and rejected for a similar application? In this case, the applicant should be advised. If the applicant wishes to test the proposed method for comparability anyway, the appropriate instructions and Data Reporting Form should be furnished to the applicant.

If the proposed method is different than approved techniques and no comparability test work has been done previously, then the applicant must submit sufficient data which will allow a statistical evaluation of comparability. Please complete lines 14 through 17 of the Preliminary Application.

If Method Comparison Tests are Required - NPDES Applications

If comparability data is required, the appropriate Detailed Equivalency Application Packet should be sent to the applicant. The waste sources to be sampled by the applicant should be listed on the front page of the instructions. In addition, one data reporting form (labeled "Comparative Test Data for Evaluating Method Equivalency") should be included for each waste source sampled. For a nationwide variance, it is recommended that samples be obtained from industries representing five different SIC code classifications. The SIC codes chosen should represent a majority of outfalls monitored for the specific parameter analyzed. It is suggested that a municipal effluent be considered as a replacement for one of the industrial effluents if the parameter is commonly monitored at municipal outfalls. In order to facilitate record-keeping, it is suggested that the EMSL application number and file number entered in line 20 of the Preliminary Application be entered in line 1 of each data reporting form sent to the applicant.

Receipt of Detailed NPDES Data Reporting Forms

Upon receipt of this form, a technical evaluation of the comparative test data can be performed. The computer program developed specifically for the statistical evaluation of this comparative test data should be used in the technical evaluation. The "Guidelines for Review of Comparative Data" may aid in the technical evaluation. As a result of the technical review, a determination regarding method equivalency can be made. This determination is made by the Deputy Assistant Administrator, Office of Research and Development, EPA for nationwide variance applications and by the Regional Administrator for limited use variance applications.

If Method Comparison Tests are Required - NIPDWR Applications

If comparability data is required, the appropriate Detailed Equivalency Application Packet should be sent to the applicant. The water sources to be sampled by the applicant should be listed on the front page of the instructions. In addition, one data reporting form (labeled "Comparative Test Data for Evaluating Method Equivalency") should be included for each waste source sampled. For a nationwide variance, it is recommended that the water sources selected be geographically distributed and represent different types of water (e.g., ground water, surface water, etc.) having a variety of chemical characteristics. In order to facilitate record-keeping, it is suggested that the EMSL application and file numbers entered in line 20 of the Preliminary Application be entered in line 1 of each data reporting form sent to the applicant.

Receipt of Detailed NIPDWR Data Reporting Forms

Completed data reporting forms will be submitted to EMSL-Ci for technical review. When the technical evaluation is completed, the recommendation concerning the application will be submitted to the Regional Administrator and the Deputy Assistant Administrator, Office of Drinking Water (DAA/ODW). Refer to the memorandum from Victor J. Kimm (op. cit.) regarding the interrelationships between the state agencies, Regional Administrators, EMSL-Ci and the office of Water Supply and the time constraints for the various steps in the review of an NIPDWR application.

APPENDIX F

REVIEWERS GUIDELINES

GUIDELINES FOR THE REVIEW OF COMPARATIVE TEST DATA

This information is intended as a guide to the reviewer in the evaluation of comparative test data submitted for an application for the use of an alternative analytical method.

Check Data Report Forms

1. Check to see that the waste or water sources tested correspond to the sources specified on the first page of the instructions. There should be one completed Data Report for each source.

2. If the application is for a nationwide variance, check whether the analytical work was performed by two or more laboratories. This is a requirement for a nationwide variance.

3. Compare all the data report sheets to be sure that the comparative testing was consistent for all sample sources. In particular, note whether the same approved procedure and preservative methods were used in all the tests. Variations in these factors could lead to inconsistencies in comparing the data.

4. Check to see whether the same units are used for all the data reporting sheets.

5. Scan the analytical data for obvious typographical errors or obvious errors in recording the data. In particular, check the contents of columns 5, 6 and 7 for format.

6. Any questions concerning the reported data can be quickly checked by calling the individual listed in line 6 of the data report form.

Guide for Review of Statistical Results

1. Compare the Computer Listing of Input Data (section 3) to the original Data Reporting Forms to check for possible key-punch errors.

2. The table of Approved and Proposed Sample Means (section 4) lists the mean value of the duplicate analyses performed on each sample

by each method.

3. The relative precision of the two methods is indicated in section 6 by using a paired t test to compare the range (difference) between the duplicate analyses for each method. If the caluclated t value is less than the "table" t value, the precision of the two methods can be considered to be comparable. If the calculated t value is greater than the "table" t, one method can be considered to be more precise than the other at the 5% significance level. A message is printed indicating whether the precision of the two methods is comparable or which of the two methods is more precise than the other.

4. The relative accuracy of the two methods is determined from the data generated by the samples prepared by spiking or by dilution. The relative accuracy of the two methods can be compared from the "% Recovery" values listed in section 7 of the computer print out.

5. For an external check on the precision and accuracy of the data from the approved method, reference can be made to the values published in different analytical procedure handbooks. There are four major sources of this type of information.

    1. Current Edition - Methods for Chemical Analyses of Water and Wastes, U.S. EPA Office of Technology Transfer, Environmental Monitoring and Support Laboratory, Cincinnati, Ohio, 46268

    2. Current Edition - Standard Methods for the Examination of Water and Wastewater, APHA, AWWA and WPCF, American Public Health Association, 1015 Eighteenth St., NW, Washington, D.C. 20036

    3. Current Edition - ASTM - Annual Book of Standards, Part 31, Water, American Society for Testing and Materials, 1916 Race Street, Philadephia, Pa. 19103

    4. Current Edition - National Handbook of Recommended Methods for Water Data Acquisition, Office of Water Data Coordination, U.S. Geological Survey, U.S. Department of the Interior, Reston, Va.

   These references can include various accuracy and precision statements for data evaluation. However, when reviewing this information, it should be remembered that analyses performed on natural samples are usually less precise than analyses performed on synthetic samples; that precision and accuracy are usually dependent on sample concentration; and that data from a single laboratory should be more precise and more accurate than data based on results from two or more laboratories. Because of these variables, the published values should be used only as rough indicators of approved method precision and accuracy for the samples analyzed by the applicant.

This type of information can also be used to check whether the approved method used by the applicant was the most appropriate method for the range of concentrations tested.

6. The skewness and coefficient of kurtosis for the reported data are used to determine whether the data are normally distributed (section 8). If the calculated values are less than the critical values, the distributions are considered normal at the 1% significance level. If one or both of the distributions are not normal, the program will attempt to make the data normal using a log, a reciprocal, or a square root transform. If none of these transforms are successful in making both distributions normal, the program will stop and inform the data analyst of the reason. If the reviewer wished, the program may be re-run and completed using the untransformed data in spite of the distribution being non-normal. A message will be printed informing the reader that the analysis was performed on data that was determined to be non-normal at a 1% significance level.

7. The paired t test is one test used to compare tne results of the two methods. If the calculated t value is less than the "table" t value, the methods can be considered to give comparable results at the 5% significance level. If the results are not comparable, a message will be printed telling which method produced the higher values. Although the paired t test is a good test for comparing results from two methods, the information available to the reviewer, when the test shows that the methods are not comparable, is very limited. The regression analysis gives the reviewer more information regarding the effects of strata, linearity, etc. on method comparability.

8. The regression analysis is also used to compare the results of the two methods. To be considered comparable by regression analysis, three conditions must be met:

   a) The data should be free of auto-correlation as indicated by the Durbin-Watson d-statistic at the 1% significance level.

   b) The slope of the regression line should not be statistically different from 1.0 at the 5% significance level, and

   c) The y-intercept of the regression line should not be statistically different from zero at the 5% significance level.

   The Durbin-Watson d-statistic may vary from 0 to 4. If there is no auto-correlation, the d-statistic will be 2.0. When the calculated value is outside the critical limits calculated for the system, this is an indication that the use of regression analysis is not appropriate for the data submitted. The regression analysis will continue, however, and the statistics will be printed. If all three conditions are met, the program will indicate that the methods can be considered comparable by the regression analysis.

9. The Durbin-Watson analysis in Section 10 of the computer print out gives the calculated d-statistic and the critical values for comparison. In addition, a table of the sample means listed in order of smallest to largest value is given. This table is easily scanned to determine whether the samples covered the concentration range of the methods computer tested.

10. The "R SQUARED" value listed in Section 10 of the computer print out is the index of determination. This value indicates the fraction of change in the y value (results of proposed method) that is associated with a unit change in the x value (result of approved method). Ideally, "R SQUARED" should be 1.0000. A value much lower than 1.0 is also an indication that the two methods do not produce comparable results.

11. If the slope of the regression line is significantly lower than 1.0, the proposed method gives lower results than the approved method. If the slope is significantly higher than 1.0, the proposed method gives higher results than the approved method.

12. If the y intercept is significantly different from zero and the slope is statistically equal to 1.0, one of the methods is affected by a constant error. If the intercept is significantly different from zero and the slope is significantly different from 1.0, there is an apparent systematic bias to one of the methods. Comparison of the accuracy data in section 7 may indicate which method is biased.

13. When evaluating equivalency of methods, special consideration should be given to the concentration range where comparability of data is most critical, such as, at or near the concentration limits specified for NPDES permits or at the maximum contaminant level for the NIPDWR.

14. A scatter plot of the data along with the calculated regression line and the ideal fit line can be extremely helpful in reviewing the data. Crossover effects, systematic bias and non-litearity are readily apparent.

15. Inspection of the scatter plot may indicate that the methods may not be comparable over the entire concentration range tested. To test method comparability over a narrower range, re-run the program using only the data from the samples covering the range of interest. The computer program is not restricted to accepting data from only 12 samples.

16. Throughout this program, the calculated t values are compared to critical values labeled as the "critical" or "table" value. These values may differ slightly from the table values published in the literature. The "table" values in this program are calculated from an equation used to approximate the published table values.

17. When evaluating comparative data from more than one sample source, it is recommended that the sources be evaluated independently and

as a group. The individual source evaluation may aid in the identification of interference or matrix effects on method comparability for certain types of industrial effluents.

18. In some cases, the paired t test and the regression analysis may give different results regarding comparability. Inspection of the data from the regression analysis and the scatter plot may aid in determining the reason for the discrepancy.